AD-A032 154 NAVAL RESEARCH LAB WASHINGTON D C F/G 17/2.1
A SURVEY OF PACKET SWITCHING TECHNIQUES FOR BROADCAST MEDIA.(U)
OCT 76 C L HEITMEYER, J H KULLBACK, J E SHORE
UNCLASSIFIED NRL-8035 NL

| OF |
AD
A032154

END
DATE
FILMED
1 — 77

NRL Report 8035

(12)

A Survey of Packet Switching Techniques for Broadcast Media

CONSTANCE L. HEITMEYER, JOSEPH H. KULLBACK,
AND JOHN E. SHORE

Information Systems Staff
Communication Sciences Division

October 12, 1976

NAVAL RESEARCH LABORATORY
Washington, D.C.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 14 NRL [redacted] 8035	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) 6 A SURVEY OF PACKET SWITCHING TECHNIQUES FOR BROADCAST MEDIA.	5. TYPE OF REPORT & PERIOD COVERED An interim report on a continuing NRL Problem.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 Constance L. Heitmeyer, Joseph H. Kullback John E. Shore	8. CONTRACT OR GRANT NUMBER(s) 9 Interim rept.,
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, D. C. 20375	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 16 Program Element 65866N Project Number XCC59 NRL Problem R01-85B
11. CONTROLLING OFFICE NAME AND ADDRESS Department of Defense Naval Electronics Systems Command Washington, D.C. 20360	12. REPORT DATE 11 12 Oct [redacted] 76
	13. NUMBER OF PAGES 68
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 12 69 p.	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribtution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Channel management
Packet switching
Satellite channels
Statistical multiplexing
Demand access
Broadcast channels

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

To utilize communications channels effectively, appropriate channel management schemes must be implemented. The Navy is considering several alternative multiplexing techniques for the future Naval Telecommunications System. This report surveys a number of recently proposed techniques, called packet switching techniques, that have been designed to multiplex broadcast radio channels. Included in the report are several random access techniques such as pure and slotted ALOHA, as well as a number of packet reservation schemes. We describe how each technique works and report the results of analytic and simulation studies of the technique's →

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251 950

20

performance. Some preliminary conclusions are stated. A subsequent report will discuss the attractiveness of these techniques for future Naval communications systems.

CONTENTS

A SURVEY OF PACKET SWITCHING TECHNIQUES FOR BROADCAST MEDIA

1.0 INTRODUCTION

Future Naval communications systems will use satellite channels extensively. Appropriate channel management techniques must be implemented to utilize these channels effectively. Circuit switching, in which the channel is dedicated to a particular conversation until that conversation is terminated, is a channel management technique under serious consideration by the Navy. One possible alternative to circuit switching is packet switching, which has been developed over the past few years as an alternative technology to circuit switching for computer communications applications. The purpose of this report is to survey those packet switching techniques that are suitable for radio channels. A subsequent report will analyze the attractiveness of these techniques for use in Naval communications systems.

In Section 2 we mention briefly the various techniques that have been proposed for allocating channels in a multi-access communications system. We then concentrate on packet switching techniques and review the major considerations that led to their development. In Section 3 we distinguish between point-to-point channels and radio channels and introduce a class of packet switching techniques whose development was stimulated by the special capabilities of radio channels. Section 3 also includes a brief summary of each technique and introduces the terminology that will prevail throughout the rest of the report. Sections 4-12 contain more detailed discussions of the techniques.* For each technique, we include a description of the channel management algorithm and a summary of what is known about the technique's performance as the result of analysis or simulation. Finally, in Section 13, we state some preliminary conclusions concerning the possible utility to the Navy of the various packet switching techniques. A more detailed analysis of this question will be the subject of a subsequent report.

Readers not interested in the details of the various channel management algorithms and their performance may find it helpful to skim or omit Sections 4-12.

2.0 CHANNEL MANAGEMENT FOR MULTI-ACCESS SYSTEMS

There have been many techniques proposed to handle the problem of how to allocate a communications channel when there are competing demands for service among the users. One alternative already mentioned is circuit switching. Another is to use a conventional fixed orthogonal multiplexing scheme such as Time-Division Multiple-Access

*Several of the techniques described are also discussed by Kleinrock [Kleinrock, 1976].

Manuscript submitted July 7, 1976.

(TDMA) or Frequency-Division Multiple-Access (FDMA). Under such a scheme, each user in the network is assigned a fixed portion of the channel. When a user is idle, however, his portion of the channel cannot be used by other stations with traffic; this results in wasted channel capacity. Moreover, like circuit-switched systems, such techniques are especially inefficient when the traffic from each user is short and bursty.

To overcome this problem, techniques have been developed that attempt to perform "statistical load averaging" of the user traffic onto the common access channel. The concept of statistical load averaging is that by assigning the channel dynamically on the basis of user needs, the channel capacity required may be much less than when the channels are dedicated and not shared. The fundamental procedure is to switch the communications channel from user A to user B when A is idle and B has something to transmit. With this type of procedure, each user is granted channel access only when he has data to transmit. One example of such a technique is a polling system [Konheim, 1974]. Another is packet switching.

Historically, packet switching was developed as an alternative to circuit switching in computer communications systems. In computer communications, the traffic is typically "bursty"; i.e., the ratio of the peak data rate to the average data rate is very large. Given the bursty nature of the traffic, dedicating a circuit-switched channel for an entire computer session represents very wasteful usage of a communications resource. An alternative is to establish a channel every time a message is exchanged between a computer and a terminal. However, the disadvantage of using circuit switching in this manner is that the connection or setup times can be prohibitively long for short, interactive messages. In a packet switched system, messages are packaged into one or more fixed-length units of information called packets.* In addition to data bits, each packet has a header which contains control information, i.e., source address, destination address, sequence number, etc.† Moreover, each packet may also include a number of parity bits for the purposes of error detection.

An important distinction between packets and messages should be noted. While messages are units of information recognized by users of the network and hence may be of variable length, packets are meaningful only within the network. Because each packet contains its own control information, there are no lengthy connection or setup times in packet-switched systems. Thus, packet switching makes efficient use of the communications channel when messages are short. Moreover, because the communications channel is dedicated to packets, not conversations, the channel can be used to support other conversations when a given conversation is inactive. Thus, packet switching makes efficient use of the communications channel when messages are bursty [Metcalfe, 1973].

Early development of packet switching concepts was based on the assumption of point-to-point communications media. In a network based on point-to-point media, as the number of nodes increases, it becomes economically infeasible to provide a fully connected

---

*While this report is concerned with fixed-length packets, packets of variable length have been investigated elsewhere. For example, see [Ferguson, 1975].

†The term "packet" is also used later in the report to refer to blocks which contain only data and no control information. In this section, we have adopted a more restrictive definition of the term in order to contrast packet switching with other allocation strategies.

network topology. Thus, partially connected networks such as the Advanced Research Projects Agency (ARPA) Network, which uses store-and-forward techniques, were designed and implemented. In the design of a store-and-forward network, however, complex problems arise with respect to network topologies, line capacities, and routing strategies.

3.0 PACKET SWITCHING FOR BROADCAST NETWORKS

In contrast to point-to-point channels, radio channels (both satellite and ground radio channels) have two important capabilities: a *broadcast* capability and a *multi-access* capability [Lam, 1974]. The application of packet switching concepts to radio channels, with their special capabilities, has led to the development of several new channel management schemes which are the subject of this report.

The *broadcast* capability of a radio channel derives from the fact that a signal generated by a radio transmitter may be received over a wide area by any number of receivers. The *multi-access* capability of a radio channel is that any number of users may transmit over the same channel. Hence, in the case of a ground radio channel, all users within line-of-sight of one another form a network that is completely connected, independent of the number of users. A similar situation exists with a satellite channel. A satellite transponder in geosynchronous orbit with the earth is a radio repeater. A signal transmitted at one frequency by a user station is received by the transponder and transmitted back to earth at another frequency. All of the user stations covered by the transponder beam make up a fully connected network.

A communications network in which all users share a broadcast channel is referred to as a broadcast network. Note that the fully connected network provided by a radio channel eliminates complex network topological problems; moreover, there is no need for complicated routing strategies, such as those required in a store-and-forward network like the ARPA Network.

The two types of broadcast media of interest in this report, satellite and ground radio channels, have a significant distinction in terms of propagation delay. For a satellite in geosynchronous orbit, the round-trip propagation delay of a transmitted signal is approximately 0.27 s, but the delay of a signal transmitted over a ground radio channel is much shorter, on the order of microseconds.

3.1 The ALOHA System

In 1970, Abramson proposed a novel multiplexing technique [Abramson, 1970] which has become known as ALOHA random access. It is this ALOHA protocol that is the forerunner of the various techniques described in this report. Under the technique, a user with a data packet simply transmits it into the channel, completely unsynchronized with the transmissions of other users. The University of Hawaii's ALOHA System [Abramson, 1973b; Binder, 1975b], developed by Abramson and others, is a broadcast network that is based on the pure ALOHA multiplexing technique.

In the ALOHA System, all users are terminals transmitting data to a central computer at the University of Hawaii; the terminals share a common UHF radio channel. A second

UHF channel is used by the central computer to transmit data and acknowledgments back to the terminals. Each terminal has buffer space for exactly one packet. As stated above, under the random access protocol, a terminal with a ready packet simply transmits into the common access channel. The terminal then initiates a time-out and waits for an acknowledgment from the central computer indicating correct receipt of the packet. If, at the expiration of the time-out, an acknowledgment has not been received, the terminal retransmits the packet. This process continues until the packet is correctly received by the central computer and an acknowledgment is received by the terminal, or until the process is stopped by the terminal.

Given the unsynchronized mode of channel access, it is possible for two or more packet transmissions to overlap in time and thus collide, destroying one another. This is detected at the central site by means of a faulty checksum in at least one of the packets. In this case, no acknowledgments are sent, and the terminals whose packets collided automatically retransmit the packets.

An important parameter in the design of an ALOHA system is the retransmission interval. If all stations that suffer a transmission conflict wait an equal amount of time before retransmitting, their transmissions will, with certainty, conflict again. Hence, it is necessary to vary the length of the retransmission interval used by each of the transmitting stations. In the University of Hawaii implementation, each terminal is assigned a fixed but unique retransmission interval [Binder, 1975b]. If the number of stations is quite large, however, such a scheme may be impractical, since those stations with large retransmission intervals will consistently experience longer delays.

An alternative policy is to use randomized retransmission delays. Each time a packet retransmission is required, the terminal involved chooses the retransmission delay from some probability distribution.

3.2 Performance of Packet-Switched Broadcast Networks

The goal of packet-switched systems such as ALOHA is to provide better channel utilization than that possible under alternative multiplexing schemes and, at the same time, to minimize transmission delay. Hence, performance evaluation of these systems is based on two performance measures, channel capacity and packet delay.

3.2.1 Channel Capacity

Given fixed-length packets, the packet transmission time T is determined by the channel transmission rate. The channel input rate S is defined as the average number of new packets generated per transmission interval T. Under steady state conditions, S is also equal to the channel throughput rate. If S_{max} is the maximum achievable value of S, then S_{max} represents the channel capacity. If it were possible to perfectly schedule packets for transmission on the channel so that no gaps or overlaps existed, a channel capacity of one packet per transmission interval would be achievable; thus S_{max} is also referred to as the channel utilization. Note that because of the interference problem in the random access schemes, the channel capacity of such schemes is always less than one.

The actual traffic on the channel consists of both newly generated packets as well as retransmissions of previously collided packets. Let G represent the average number of packets per transmission interval from both sources. Then, G is called the average channel traffic. Note that the ratio S/G is the probability of a successful packet transmission.

3.2.2 Packet Delay

Another performance measure of interest in an ALOHA system is the average delay incurred by a packet before successful transmission. The delay is a function of two parameters; one is the average number of times a packet requires retransmission, while the second is the average retransmission delay.

3.2.3 The Problem of Instability and Saturation

Unfortunately, an ALOHA channel is vulnerable to unstable behavior and channel saturation. A channel operated in ALOHA mode stays in a quasi-stationary state for a finite time, until stochastic fluctuations produce an increase in the traffic rate. This increased traffic rate causes an increase in packet collisions which in turn produces an even higher traffic rate. This vicious cycle continues until the channel is filled with retransmissions and packet collisions, and the channel throughput approaches zero. At this stage, the channel is said to be *saturated.*

3.3 Other ALOHA Random Access Schemes

A number of variations of pure ALOHA have been proposed. Several of these—slotted ALOHA, slotted ALOHA with dynamic channel control, and Carrier Sense—are briefly described in this section with a more detailed discussion of each of these techniques to follow in Sections 4-7. All of these schemes have been designed to handle traffic consisting of one-packet messages.

3.3.1 Slotted ALOHA Random Access

Recall that in a pure ALOHA channel all packet transmissions are completely unsynchronized. Given a packet transmission time of T seconds, a given packet is vulnerable to collisions for a period of $2T$ seconds. If channel time is divided into contiguous intervals of length T, and if each packet transmission is required to coincide with an interval, the vulnerable time for a packet can be reduced to T seconds. A protocol identical to pure ALOHA, but with transmissions constrained in this manner, is called slotted ALOHA random access [Kleinrock and Lam, 1973]. The channel capacity of a slotted ALOHA channel is twice that of pure ALOHA, but this increase in channel capacity is realized at the cost of providing a global clock for synchronization.

3.3.2 Slotted ALOHA with Dynamic Control Procedures

Both the pure ALOHA and slotted ALOHA techniques are vulnerable to unstable behavior and saturation. To deal with the instability problem under temporary overload conditions, the use of dynamic control procedures [Lam, 1974] has been proposed.

These require each user to take action to prevent channel saturation. The broadcast nature of the communications channel permits all users to monitor the state of the channel. When the number of users with backlogged packets (i.e., packets that have suffered a collision) exceeds a given threshold, each user is required to take action to reduce the channel traffic rate and thereby prevent saturation.

3.3.3 Carrier Sense Multiple Access

The Carrier Sense technique [Tobagi, 1974] takes advantage of the broadcast nature of a radio channel to reduce the probability of a packet collision. A user with a packet to transmit first listens or "senses" the channel to determine whether another user's carrier is present. If such a carrier is present, the channel is assumed to be busy, and the user with the ready packet delays transmission until the channel is idle (i.e., no carrier is sensed). Using one of the Carrier Sense protocols, channel utilization of approximately 80% is possible. Unlike the other techniques discussed in the report, which apply to both satellite and ground radio channels, the use of Carrier Sense is most appropriate with ground radio systems.

3.4 Reservation Techniques

While the ALOHA techniques described above are most applicable to traffic composed of one-packet messages, a number of techniques called reservation schemes have been developed for traffic that contains a significant portion of multi-packet messages. Rather than make an access request for each packet, one request is made per message, where a message consists of one or more packets. Several reservation schemes are noted in this section. A more detailed discussion of each of these techniques can be found in Sections 8-12.

3.4.1 Roberts' Reservation

Under this protocol [Roberts, 1973], the channel is divided into two subchannels; one, operated in slotted ALOHA mode, is for reservation requests while the other, operated in a dedicated mode, is for data packets. The reservation requests are for the data packet slots. Because of the broadcast nature of the channel, all users can hear the successful requests for data slots. (Since the requests are made in slotted ALOHA mode, there will be some collisions, and hence not all requests will be transmitted successfully.) Based on the reservations made by other users and its own successful requests, a given user can schedule the transmission of its packets on the data channel.

3.4.2 Split Channel Reservation Multiple Access

This technique [Tobagi, 1974] is similar to the previous one in that the channel is split into two subchannels, one for control information and one for data. The control channel is accessed in random access mode; pure ALOHA, slotted ALOHA, and Carrier Sense are all alternative schemes for handling this channel. This technique differs from Roberts' Reservation in that there exists a central station which controls allocation of the data channel; in Roberts' scheme, control is distributed among all users. Under Split

Channel Reservation, a user with data sends a transmission request via the control channel to the central station. The central station then reserves the data channel for the requesting user and tells the user when to begin transmission of the data.

3.4.3 Reservation-ALOHA

This technique [Crowther, 1973] is based on a slotted ALOHA channel. In addition, a fixed number of slots are grouped together to form a frame; this imposes a periodicity on the slots similar to that in TDMA. However, unlike TDMA, in Reservation-ALOHA no slot is permanently assigned to a user. This protocol is dependent on the broadcast nature of the channel, which permits all users to monitor channel activity. A user with traffic transmits a packet into a slot which was emtpy in the previous frame. If the transmission is successful (i.e., does not collide with another user's attempted transmission), the user "owns" that slot for as long as it has packets to transmit. Thus, once a user gains access to a particular slot in a frame, no other user may transmit in the corresponding slot in subsequent frames until the owner relinquishes the slot by ceasing to transmit.

3.4.4 Round Robin Reservation

This access technique [Binder, 1975a], like the previous one, uses a slotted channel with a frame structure. In addition, each user is permanently assigned a particular slot in each frame. Thus, the basis of the protocol is a fixed TDMA structure. Moreover, a second feature is included permitting other users to transmit in slots which belong to a user who currently has no traffic to send. This is accomplished by using a distributed queue, whose contents are known to all users due to the broadcast nature of the channel. Neither a central control station nor a separate control channel is needed; the necessary control information is carried as part of each packet's overhead. This multiplexing technique combines the basic fairness and stability of fixed TDMA with the dynamic capability of the ALOHA schemes.

3.4.5 Conflict Free Multiple Access

This multiplexing technique [Hwa, 1975], very similar to the Round Robin scheme, uses a fixed TDMA structure with a control mechanism somewhat different from Round Robin to dynamically allocate currently available slots. As the name implies, under this protocol, no conflicts are ever generated.

3.5 General Remarks

Starting with pure ALOHA and moving on to the subsequent schemes, the reader should note that the latter schemes have been designed to overcome the two serious disadvantages (for some applications) of pure ALOHA; i.e., the low channel capacity and the vulnerability to unstable behavior. The newer techniques attempt to minimize collisions and control stability at the expense of channel management algorithms that are more complicated than those of pure ALOHA and, therefore, more expensive to implement.

Each of the foregoing techniques is reviewed in more detail in the following sections. The presentation for each technique consists of a description of how the technique

works, a summary of the available mathematical analysis, a description of simulation results, and a concluding discussion. We have attempted to keep each presentation reasonably self-contained. This leads to some redundancy.

The material presented in the next sections does not cover every packet switching technique that has been proposed for broadcast media, nor have we included all analytic and simulation results for those techniques included in the report. Our goal instead has been to present material that is representative of this class of techniques so as to provide insight into what can and what cannot be done with such procedures. Further details on each technique can be found in the references cited.

4.0 PURE ALOHA RANDOM ACCESS

4.1 Description

The application of packet switching concepts to broadcast communications media led in 1970 to the development of the technique known as pure ALOHA random access. First described [Abramson, 1970] by Dr. Norman Abramson of the University of Hawaii, this scheme is designed to permit a large number of stations to communicate via a single ground radio channel. However, as Abramson [Abramson, 1973a] and others have indicated, this technique and variations of it are also attractive for multiplexing a satellite channel.

Under pure ALOHA, information is transmitted in the form of packets; typically, each packet has a fixed length. The ALOHA technique allows several stations to share a single communications channel. However, there is no central control over the channel, nor is there any synchronization among the stations which share it. Channel access is gained on a contention basis. If a station has traffic, the station transmits immediately, without any coordination with the other stations; moreover, rather than use only a part of the channel capacity, the station utilizes all of the available bandwidth. If, during the packet transmission time, no other stations transmit, the transmission is successful.* If, however, two or more stations attempt transmission at the same time, none of the transmissions succeeds, and each station must retransmit at some future time. See Figure 4.1 for an example of how pure ALOHA works.

There are two ways to detect the success or failure of a transmitted packet. One way is to design each station so that it can hear its own transmission; if the transmitting station receives its own packet correctly, it assumes that no conflict with other packets occurred, and hence, that the packet transmission was successful. An alternative scheme, based on the use of positive acknowledgments, is used in the implementation of random access ALOHA at the University of Hawaii. A station, upon transmission of a packet, initiates a time-out; if, at the expiration of the time-out, no acknowledgment is received from the station to which the packet was sent, the originating station retransmits the packet. (In the Hawaii implementation, acknowledgments are sent over a second broadcast channel.)

---

*We are ignoring random noise errors.

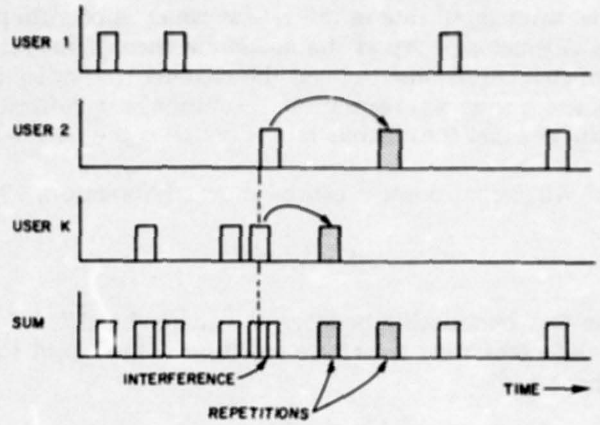


Fig. 4.1—Pure ALOHA Random Access: an example [Abramson, 1973b, p. 510]*

*This notation is used to indicate the source for a given figure.

An important parameter in the design of a pure ALOHA system is the retransmission interval. In the University of Hawaii implementation, each station is assigned a fixed but unique retransmission interval. If the number of stations is quite large, such a scheme may be impractical, since those stations with large retransmission intervals will consistently experience longer delays. An alternative is for each station to randomize the length of the retransmission interval each time a packet transmission is required.

4.2 Analysis

The model of station traffic assumed in the analysis of the pure ALOHA technique is called the infinite population model, which consists of a large number of "small" users. A small user is defined as one whose average data rate is small relative to the channel transmission rate. The combined input of all users constitutes the input source to the communications channel. In our review of the analysis, it is useful to distinguish between two kinds of packets: a) "new" packets, and b) "repetitions" or retransmissions of packets which suffered a collision.

Consider S, the average rate at which the input source generates new packets, and G, the average rate at which both new packets and repetitions are transmitted over the channel. G is called the average channel traffic. The following assumptions are made about the channel input process and the channel traffic process:

(a) Each is an independent process.

(b) Each is Poisson-distributed.

(c) Each has a stationary probability distribution.

We define the channel throughput rate as the rate at which successful packet transmissions are received and the channel capacity as the maximum throughput rate. To find the relationship between the channel throughput and the channel traffic, equilibrium conditions are assumed (assumption (c) above). Equilibrium solutions are defined as those values of S and G such that the channel throughput rate is equal to the channel input rate.

Based on the above assumptions, it can be shown [Abramson, 1973 a & b] that

$$S = Ge^{-2G}.$$

Figure 4.2 illustrates this relationship between the channel traffic rate and the channel throughput rate. By differentiating the above equation with respect to G, we can show that the channel capacity is

$$S_{max} = \frac{1}{2e} \approx 0.184.$$

Thus, under pure ALOHA, channel utilization is restricted to about 18%.

The analytic results presented above are based on the assumption of equilibrium conditions. However, as pointed out by Lam [Lam, 1974] in his study of slotted ALOHA (a variation of pure ALOHA), this assumption may not be valid. Because of stochastic fluctuations in the channel input, channel saturation may occur; i.e., an increase in new arrivals may decrease the channel throughput, which, in turn, produces an increase in the channel traffic. Rapidly, the channel is filled with collisions and retransmitted packets, and the channel throughput vanishes to zero. This instability problem is further explored in Section 5.

Unlike slotted ALOHA whose performance has been studied extensively (see Lam [Lam, 1974], for example), the performance analysis of pure ALOHA completed to date is still quite limited. The analysis of the delay performance of this technique as well as its time-dependent behavior have not been explored.*

4.3 Simulation

Simulation has produced excellent agreement, with the analysis for S, the average input rate, less than 0.15 [Bortels, 1970]. For larger values of S, the system becomes unstable.

4.4 Discussion

In contrast to more conventional multiplexing techniques, the pure ALOHA scheme is attractive because of its inherent simplicity. Moreover, the cost of implementing this

*Recently, a report [Kobayashi, 1976] has become available which presents major new results for the performance of pure ALOHA. Unfortunately, time considerations preclude the inclusion of these results here.

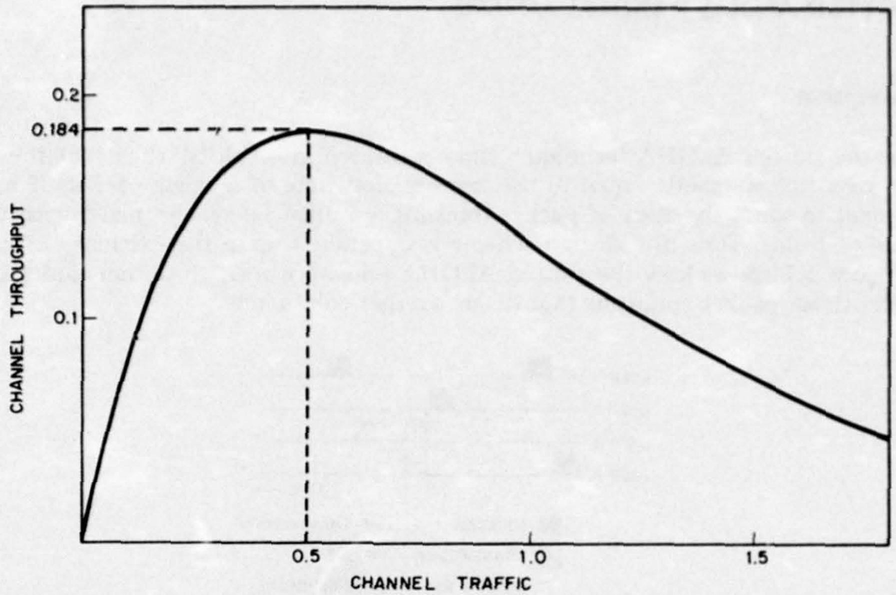


Fig. 4.2—Channel traffic vs channel throughput for a pure ALOHA channel

scheme is low relative to the implementation costs of alternative techniques. An additional advantage of the ALOHA protocol is that, in contrast to synchronous time-division and frequency-division multiplexing systems, an ALOHA station uses no part of the channel capacity when it is idle; when the station has traffic, however, its transmission may utilize the entire bandwidth of the channel.

A discouraging aspect of the ALOHA technique is its low channel capacity; the maximum channel utilization is less than one-fifth. An even more serious problem in some applications is the inherent instability of such a system. Unfortunately, there is no mechanism in the pure ALOHA protocol to determine when the system is operating near saturation. Hence, it is not possible to take action to prevent the system from saturating.

One variant of the pure ALOHA scheme, called ALOHA with capture, has been proposed by Roberts [Roberts, 1972]. A characteristic of radio receivers is that they can receive several simultaneous transmissions and capture only one of them if the power of that transmission is sufficiently stronger than the power of the others. This capture effect can be used in a pure ALOHA system so that a packet collision need not prove fatal to all of the packets involved. Thus, under a system with capture, one of the packets involved in a collision may be received correctly. The result is a higher theoretical channel capacity for an ALOHA system with the capture feature than one without it. For further details see Roberts [Roberts, 1972].

5.0 SLOTTED ALOHA RANDOM ACCESS

5.1 Description

For the slotted ALOHA technique, time is divided into "slots" such that the duration of a slot is exactly equal to the transmission time of a single packet. If a station has a packet to send, the start of packet transmission must be synchronized with the beginning of a slot. Thus, the slotted scheme is dependent upon the existence of a global clock. Figure 5.1 shows how the slotted ALOHA scheme works. Note that under this technique, those packet collisions that occur overlap completely.

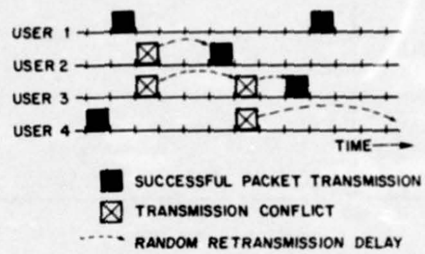


Fig. 5.1—Slotted ALOHA Random Access: an example [Lam, 1974, p. 11]

5.2 Analysis

Three different user models have been assumed for the analysis of the slotted ALOHA technique: the infinite population model, the large user model, and the finite population model. Each model, along with its analytic results, is described below.

5.2.1 Infinite Population Model: Equilibrium Solution

An initial result of Roberts' analysis [Roberts, 1972] of slotted ALOHA concerns channel capacity. The analysis is based on the infinite population model described in Section 4 and uses the same three assumptions about the channel input process and the channel traffic process, i.e., Poissonness, independence, and stationarity. If S is the average input rate and G is the average traffic rate, it has been shown [Roberts, 1972] that

$$S = Ge^{-G}$$

with the maximum throughput on the channel equal to

$$S_{max} = \frac{1}{e} \cong 0.368 . \qquad (5.1)$$

Note that the theoretical capacity of slotted ALOHA, $1/e$, is twice that of pure ALOHA, $1/2e$ (Figure 5.2). An intuitive explanation for this factor-of-two increase follows: In pure ALOHA, a given packet will collide with another packet if there is a packet transmission beginning within T seconds before or after the start time of the given packet, where T is one packet transmission time. In slotted ALOHA, if two packets collide, they will overlap completely. Hence, the vulnerable period for a packet in pure ALOHA is $2T$ and for slotted ALOHA is T.

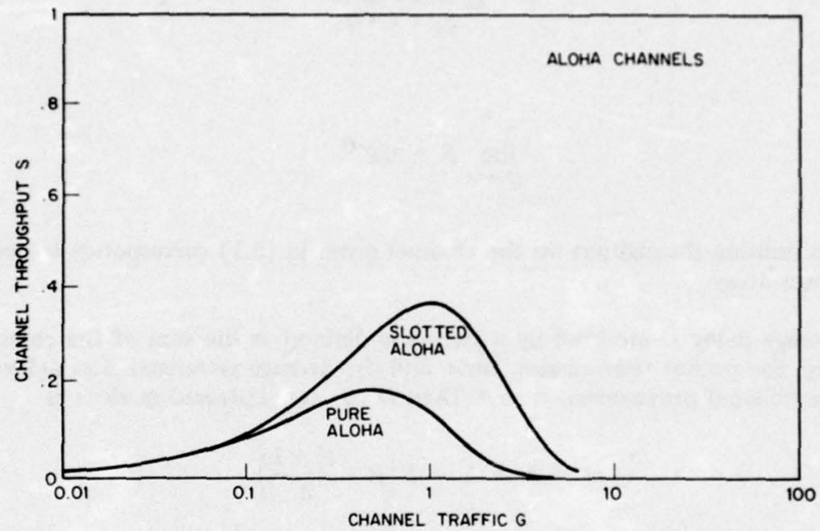


Fig. 5.2—Channel traffic vs channel throughput for pure ALOHA and slotted ALOHA [Tobagi, 1974, p. 25]

The model used in Roberts' analysis fails to distinguish between new packets and previously collided packets. A more accurate model, suggested by Lam [Lam, 1974], defines a small user as one with buffer space sufficient for storing up to one packet. Then, if the buffer is empty, the user generates a new packet with probability σ; if the buffer is not empty, the user is blocked until the buffer's contents have been successfully transmitted. The analysis by Roberts also neglects an important system parameter, retransmission delay, which is defined as the time from a packet's collision in the channel to its subsequent retransmission. By extending the analysis to include retransmission delay and by distinguishing between new and previously collided packets, it is possible to gain insight into the throughput-delay tradeoffs associated with the slotted ALOHA technique.

Assume that the retransmission delay is distributed randomly and uniformly over K slots. Then, define q_n as the probability of a successful packet transmission, given transmission of a new packet, and q_t as the probability of a successful packet transmission, given the transmission of a previously collided packet. Lam has shown [Lam, 1974] that

$$q_n = \left[e^{-G/K} + \frac{G}{K} e^{-G} \right]^K e^{-S}$$

and

$$q_t = \frac{e^{-G/K} - e^{-G}}{1 - e^{-G}} \left[e^{-G/K} + \frac{G}{K} e^{-G} \right]^{K-1} e^{-S} .$$

The relationship between the channel throughput and the channel traffic is given by

$$S = G \frac{q_t}{q_t + 1 - q_n} .$$

Note that

$$\lim_{K \to \infty} S = Ge^{-G}$$

and thus the limiting throughput on the channel given in (5.1) corresponds to the case of infinite average delay.

The average delay D incurred by a packet is defined as the sum of the channel propagation delay, the packet transmission time, and the average retransmission delays. Let R represent the channel propagation time.* Then D can be expressed in slots as

$$D = R + 1 + E\left(R + \frac{K+1}{2}\right)$$

where E is the average number of retransmissions per packet and $(R + (K + 1)/2)$ represents the average retransmission delay.

Let q represent the probability that a packet is transmitted successfully. It has been shown [Lam, 1974] that

$$q = \frac{S}{G} = \frac{q_t}{q_t + 1 - q_n} .$$

In Figure 5.3, q is plotted as a function of K, the number of retransmission slots, for several different channel traffic rates. Note that if we fix the traffic rate G, then q rapidly approaches its theoretical limit of e^{-G}. Note further that for a fixed K, q increases as G decreases.

While Figure 5.1 shows the channel throughput/channel traffic relationship for the limiting case of slotted ALOHA (i.e., for $K = \infty$), Figure 5.4 illustrates this same relationship for several finite values of K. Recall that under the assumption of equilibrium conditions, the channel throughput rate is equal to the channel input rate. If we fix G, then the throughput rate and K increase together, with the maximum throughput equal to

*For a satellite channel, R typically will have a duration of several slots, whereas for a ground radio channel, R will be restricted to a fraction of a slot. For the results given, the channel is assumed to be a satellite channel, and R has been assigned a value of 12 slots.

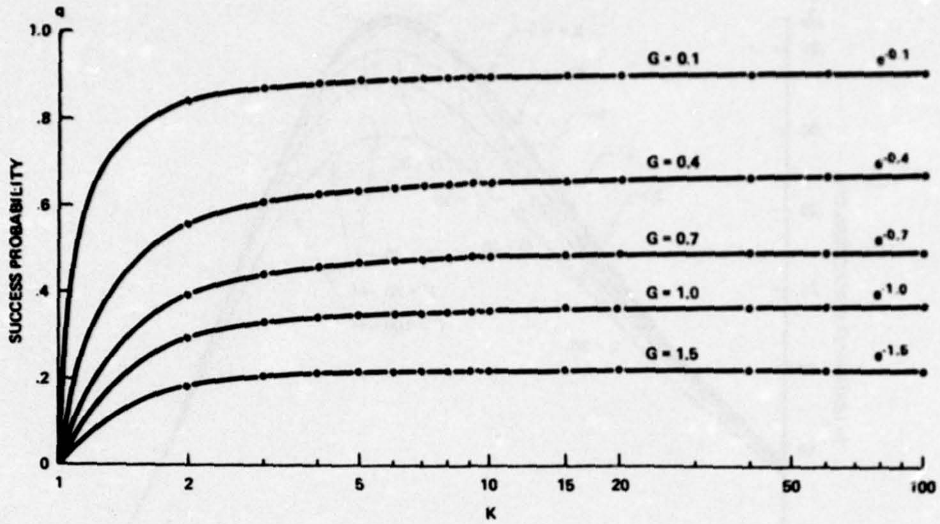


Fig. 5.3—Slotted ALOHA: probability of a successful transmission as a function of K [Lam, 1974, p. 48]

Ge^{-G} at $K = \infty$. Note that the maximum throughput occurs at $G = 1$ for each K, and channel capacity approaches e^{-1} as K approaches infinity.

In Figure 5.5, the tradeoff between throughput and delay is shown. Throughput-delay equilibrium contours are drawn for fixed values of K; the minimum envelope of these contours shows the optimum channel performance for this technique. Note that near the maximum throughput for a particular K, a small increase in throughput is accompanied by a very large increase in delay.

5.2.2 Infinite Population Model: Time-Dependent Analysis

In [Lam, 1974], an exact mathematical model of a time-dependent slotted ALOHA channel is given. Using only the independence assumption for the input process, Lam has derived a complicated transform equation to characterize the time-dependent behavior of the channel. Unfortunately, no simple solution to that equation has been found.

To obtain an approximate solution, Lam makes the further assumption that the channel traffic is independent within any K slots. He is then able to derive a difference equation which provides an approximation of the dynamic behavior of the channel subject to time-varying inputs. Lam uses this approximation to study the effect of time-varying inputs on channel stability, and concludes that the approximation results agree very well with the general trend of the simulations done on the performance of slotted ALOHA. That is, the assumption of equilibrium is valid for only a finite time period beyond which the channel saturates. Further details of the simulation results are presented in Section 5.3.

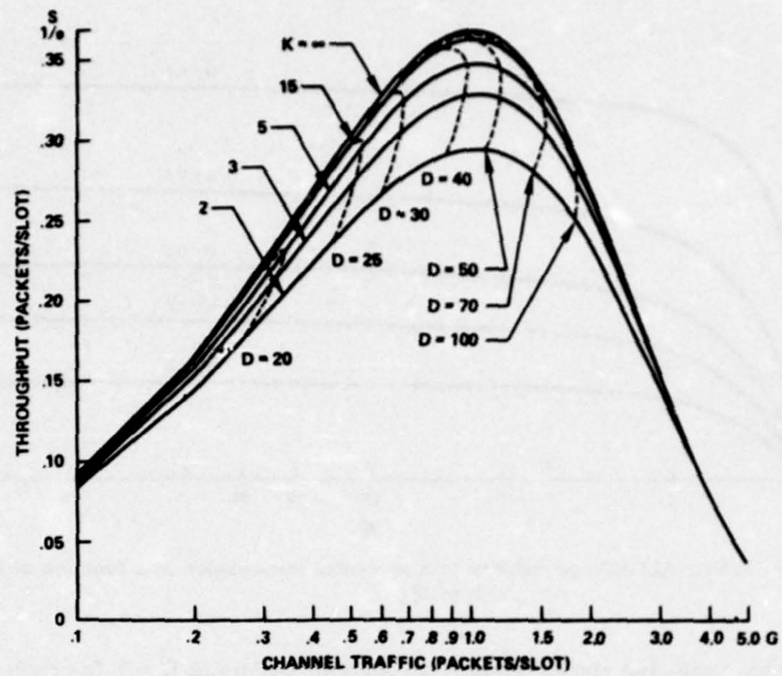


Fig. 5.4—Slotted ALOHA: channel traffic vs channel throughput for different values of K [Lam, 1974, p. 50]

5.2.3 Large User Model: Analysis

A second model studied in connection with the slotted ALOHA technique is the "large-user" model, in which there exist two different sources of traffic; one is called the background source and the other is referred to as the large-user source. The background source operates exactly like the infinite population model described earlier; in that model, the user population is composed of a large number of small users, each of whom has storage for only one packet. The second source of packets in the large-user model is a single, large user who is assumed to have infinite storage for packets as well as scheduling capability. Unlike the background source, the larger user does not attempt simultaneous transmission of packets with itself. Instead, this user can queue packets and then schedule their transmission according to some priority rule.

For a slotted ALOHA system with a fixed average input rate, one can compare the maximum throughput achievable with the large population model with that possible with the infinite population model. The channel capacity can be significantly greater for the large user model, since the large user can queue simultaneous demands from its input sources and thus reduce the number of collisions in the channel.

For the analysis of this model, let S_1 and G_1 represent the average input rate and the average traffic rate, respectively, for the background source, and let S_2 and G_2 represent the corresponding parameters for the large user. Assume that the two channel input processes—one for the collection of small users and one for the larger user—are

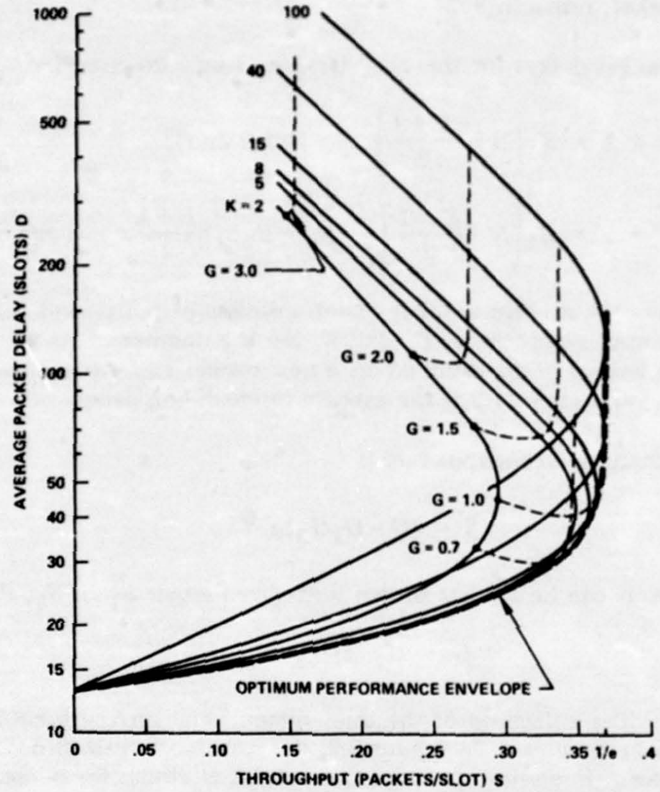


Fig. 5.5—Throughput-delay tradeoff for slotted ALOHA [Lam, 1974, p. 51]

independent, each with a stationary Poisson distribution. Make the same assumption about the two traffic processes. Let S represent the channel input rate where

$$S = S_1 + S_2$$

and let G represent the channel traffic rate with

$$G = G_1 + G_2 .$$

The equilibrium solutions for this model [Lam, 1974], are similar to those obtained for the infinite population model:

$$S_1 = G_1 \frac{q_{1t}}{q_{1t} + 1 - q_{1n}}$$

and

$$S_2 = G_2 \frac{q_{2t}}{q_{2t} + 1 - q_{2n}}$$

where q_{in} and q_{it} ($i = 1, 2$) represent the probability of success of a new packet or a previously collided packet, respectively.

The average packet delays for the two classes of users are given by

$$D_1 = R + 1 + E_1\left[R + \frac{K+1}{2}\right] \qquad \text{(small user)}$$

$$D_2 = R + 1 + E_2\left[R + \frac{K+1}{2}\right] + (E_n + E_2 E_t)\,\frac{L+1}{2} \qquad \text{(large user)}$$

where E_1 and E_2 are the average number of retransmissions per packet for the small users and the large user, respectively; E_n and E_t are the number of reschedules per packet transmission at the large user conditioned on a new packet and a previously collided packet, respectively; and $(L + 1)/2$ is the average rescheduling delay.

The limiting channel throughput rate is

$$S = (G - G_1 G_2)e^{-G_1}.$$

From this equation, it can be further shown that, given either S_1 or S_2, S is maximized if

$$G = G_1 + G_2 = 1.$$

Figure 5.6 provides a diagram of the three-dimensional surface for S as a fui ction of G_1 and G_2 for the limiting case. In Figure 5.7, the maximum throughput contour at $G_1 + G_2 = 1$ is shown. In addition, several throughput contours for constant values of G_1 are given. Figure 5.8 illustrates the throughput-delay performance at $S_1 = 0.1$, where D, the average delay on the channel, is defined as $(S_1D_1 + S_2D_2)/S$. Note that, if one is willing to drive the input rate of the small users down to 0.1, the channel throughput increases to a maximum of approximately $S = 0.52$. However, this gain in maximum throughput is accompanied by increased delays, especially for small users.

Figure 5.9 shows the optimum throughput-delay performance contours for various values of S_1. With values of S_1 less than 0.1, significant gains in maximum throughput are made. The absolute optimum channel performance is obtained when the channel is modeled as a single-server queue, i.e., when $S_1 = 0$ and only the large user is generating traffic. In this case, a channel throughput rate arbitrarily close to unity is achievable. Note that a continuum of throughput-delay tradeoff performances exist between two extremes—the single-server queuing model at one end and the infinite population model on the other.

5.2.4 Finite Population Model: Analysis

The "finite population" model is a third model for which the slotted ALOHA technique has been analyzed. Here, all users are large; hence, each user has the buffering and scheduling capabilities associated with a large user.

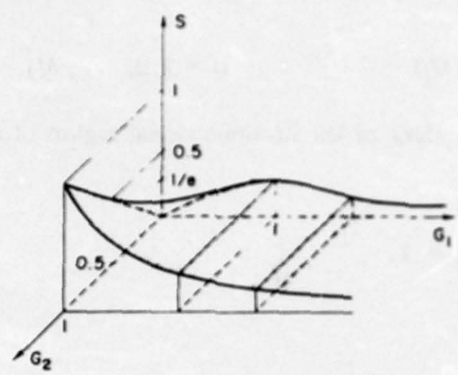


Fig. 5.6—Slotted ALOHA: throughput surface for the large user model [Lam, 1974, p. 63]

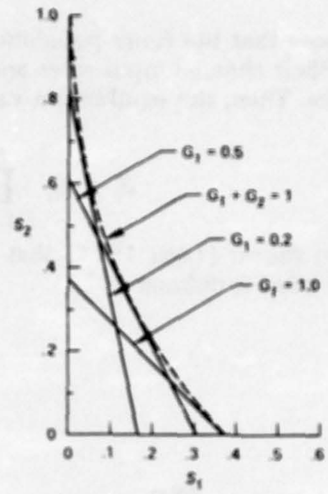


Fig. 5.7—Slotted ALOHA: allowable throughput rates for the large user model [Lam, 1974, p. 63]

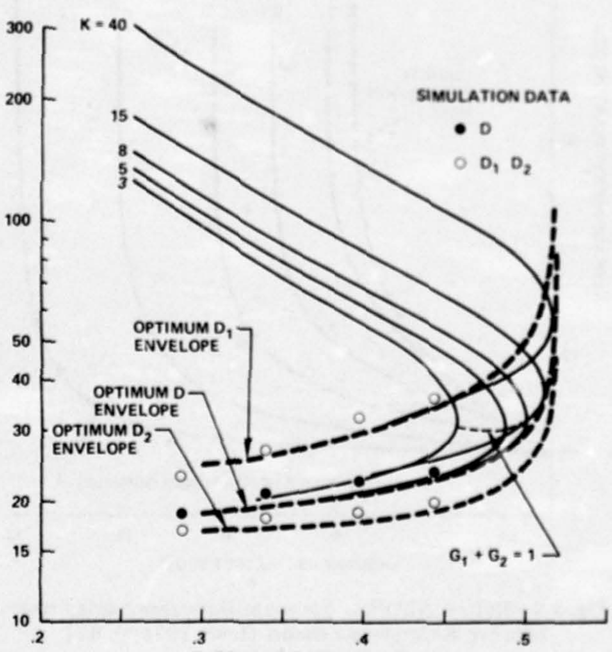


Fig. 5.8—Slotted ALOHA: throughput-delay tradeoff at $S_1 = 0.1$ for the large user model [Lam, 1974, p. 65]

Suppose that the finite population consists of M large users. Let $S_1, S_2, \ldots, S_M$ represent their channel input rates and let $G_1, G_2, \ldots, G_M$ represent their channel traffic rates. Then, the equilibrium values of S_i and G_i are given [Lam, 1974] by

$$S_i = G_i \prod_{j=1, j \neq i}^{M} (1 - G_j) \qquad (i = 1, 2, \ldots, M)$$

It has been shown [Lam, 1974] that the boundary of the M-dimensional region of allowable input rates is defined by

$$\sum_{i=1}^{M} G_i = 1.$$

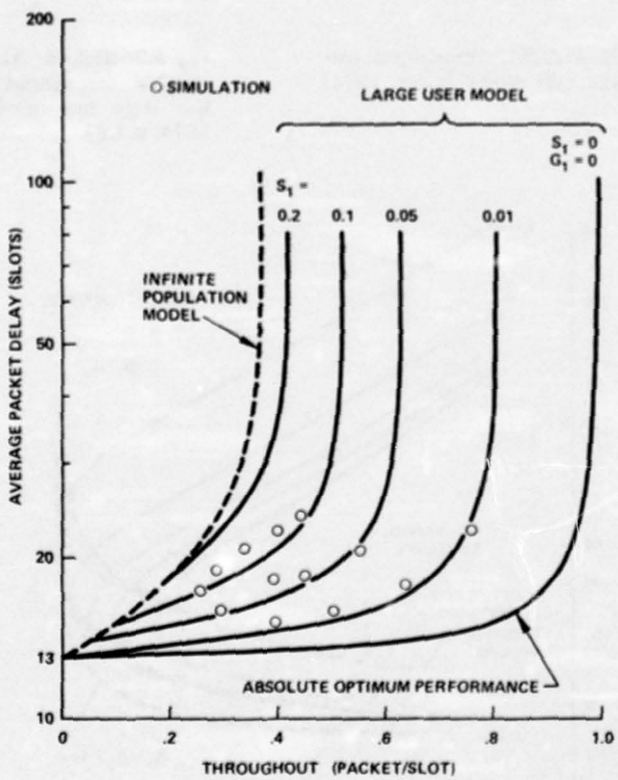


Fig. 5.9—Slotted ALOHA: optimum throughput-delay trade-offs for the large user model [Lam, 1974, p. 67]

Consider the following special case of this model. Suppose there exist two groups of users with M_1 users in group 1 and M_2 users in group 2. Suppose further that S_1/M_1 and G_1/M_1 represent the input rate and traffic rate for each user in group 1 and that S_2/M_2

and G_2/M_2 are the corresponding parameters for each user in group 2. The M equations given above become

$$S_1 = G_1\left(1 - \frac{G_1}{M_1}\right)^{M_1 - 1} \cdot \left(1 - \frac{G_2}{M_2}\right)^{M_2}$$

$$S_2 = G_2\left(1 - \frac{G_2}{M_2}\right)^{M_2 - 1} \cdot \left(1 - \frac{G_1}{M_1}\right)^{M_1}$$

and the boundary of allowable input rates is

$$G_1 + G_2 = 1 .$$

Then, maximum throughput contours for various values of M_1 and M_2 can be computed. Figure 5.10 gives several examples of these contours. Note that the special cases (∞, ∞) and $(\infty, 1)$ correspond to the infinite population model and the larger user model, respectively.

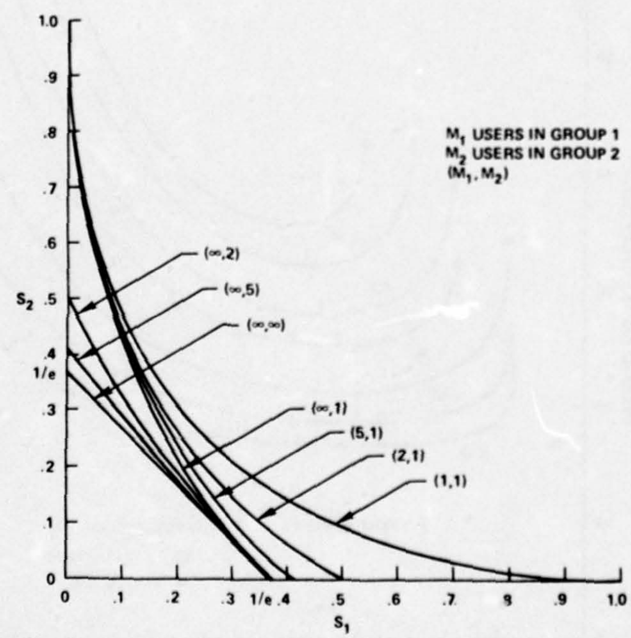


Fig. 5.10—Slotted ALOHA: Allowable throughput rates for the finite population model [Lam, 1974, p. 72]

5.3 Simulation

Simulation results for the slotted technique have been obtained for all three user models [Lam, 1974]. In each case, the analytic and simulation results agree very well, *with one important exception. For all three models, simulation has shown that the equi*librium assumption is valid for only a finite time period beyond which the channel goes into saturation. Each simulation run behaves in the following manner. Starting from an empty system, the system stays in equilibrium for a finite time period until stochastic fluctuations give rise to an increased traffic rate. This produces an increase in packet collisions which in turn causes a further increase in the traffic rate. As this vicious cycle continues, the channel is filled with collisions and retransmissions, and the channel throughput rapidly vanishes to zero. The length of time the system stays in equilibrium depends upon both S and K. As one would expect, this time period increases with a decrease in S or an increase in K. As an example, for the simulation run with $S = 0.35$ and $K = 15$, the channel stayed in equilibrium for only 3000 time slots and then rapidly saturated.

In Figure 5.11, several simulation points which show the relationship between packet delay, S, and K are given for the infinite population model. The simulation and analytic

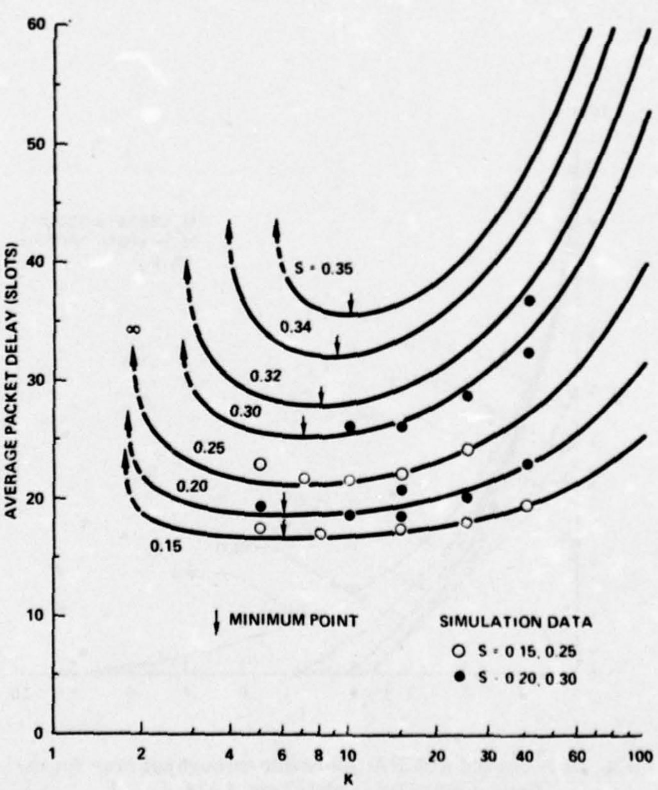


Fig. 5.11—Slotted ALOHA: average packet delay vs K [Lam, 1974, p. 52]

results agree very well; however, the highest S shown is only 0.3, since if S exceeds that figure, the channel saturates very quickly.

Figure 5.12 shows the throughput-delay performance predicted by simulation for the finite population model and populations of 2, 3, 5, and 10 users. Note that when the user population is 10, the results are very close to those of the infinite population model, and the large-user effect is minimal.

5.4 Discussion

As discussed above, using a slotted technique with the small users of the infinite population model permits a maximum throughput rate twice that associated with pure ALOHA; recall that for a large population of small users the slotted technique can theoretically support a channel utilization of up to 36%. In the large-user model, the maximum channel throughput can be increased even further.

However, the increase in channel capacity associated with slotted ALOHA is accompanied by some sacrifice with respect to the simplicity of the technique. Unlike pure ALOHA, the slotted scheme requires a global clock for synchronization of user packets into slots. Providing this synchronization is not a trivial problem.

As with the pure ALOHA scheme, slotted ALOHA has a serious problem with instability. If channel utilization is close to its theoretical capacity, the channel saturates within a short period of time. As with the pure ALOHA scheme, a user has no positive way

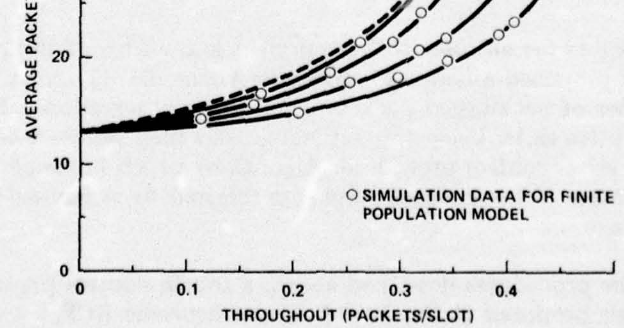


Fig. 5.12—Slotted ALOHA: Throughput-delay tradeoffs for the finite population model [Lam, 1974, p. 73]

of recognizing when the channel is operating in a dangerous zone; hence, the users cannot take action to control transmission in the channel and thereby prevent saturation.

A variant of pure ALOHA, called pure ALOHA with capture, was mentioned in Section 4.4. A slotted ALOHA system with the capture feature has also been described and analyzed. See [Roberts, 1972] for details.

6.0 SLOTTED ALOHA WITH DYNAMIC CHANNEL CONTROL

6.1 Description

Like other contention schemes, the slotted ALOHA technique (see Section 5) is vulnerable to unstable behavior and channel saturation. In applications where a minimum level of channel performance is required, such behavior is unacceptable. To deal with the problem of instability under temporary overload conditions, the use of dynamic control procedures has been proposed [Lam, 1974] (see also [Kleinrock and Lam, 1975a; Lam & Kleinrock, 1975]). These procedures are designed to convert an unstable channel into a stable one by requiring each user to take action to prevent channel saturation. The action taken may assume a variety of forms.

Three of the control procedures that have been suggested [Lam, 1974] require all users to monitor the channel's behavior; i.e., each user must keep track of recent channel transmissions (both successful and unsuccessful) as well as empty slots. When the number of backlogged users (i.e., users with packets which have suffered one or more collisions) exceeds a certain threshold value, each user is required to take action to reduce the rate of transmission attempts. The Input Control Procedure (ICP) is one type of control procedure which has been proposed. ICP requires all users to reject new packet transmission requests once the threshold value is exceeded. When the channel returns to a "safe" level, users may again accept new packets for transmission. A second procedure is the Retransmission Control Procedure (RCP), under which the number of retransmission slots K is increased as soon as the number of backlogged packets becomes excessive. When the channel activity is reduced sufficiently, K is decreased to the value which it held prior to the traffic surge. A third procedure, called the Input-Retransmission Control Procedure (IRCP), calls for the application of both ICP and RCP.

Since it is impossible for all users to have perfect knowledge of the recent history of the channel, Lam has proposed a heuristic procedure [Lam, 1974] that allows each user to estimate the number of backlogged packets based upon observations of channel activity for W consecutive slots. Using this estimate, users may decide when to take the action specified by a given control procedure. Algorithms which implement a control procedure using this estimate of the channel status are referred to as control-estimation or CONTEST algorithms.

In addition to the procedures described above, a fourth control procedure called heuristic RCP has been proposed [Lam, 1974]. Under heuristic RCP, a user with a backlogged packet uses a retransmission interval $K = K_m$, where m is the number of times the given packet has been retransmitted and K_m is a monotone nondecreasing function in

m. Thus, under heuristic RCP, as channel traffic increases, the retransmission delays of backlogged packets also increase, and the risk of channel saturation is reduced. Since it requires neither monitoring of the channel history nor estimation of the channel state, this procedure is implemented more easily than the procedures described above.

6.2 Analysis

To evaluate the effect of dynamic control procedures on an unstable channel, Lam has formulated the following model of slotted ALOHA [Lam, 1974] (see also [Kleinrock and Lam, 1975b]). A similar model has been reported by Carleial and Hellman [Carleial & Hellman, 1975]. In Lam's model, each of M users is in one of two states: thinking or blocked. In the thinking state, a user generates a new packet in a given time slot with probability σ. A user in the blocked state is one who has a backlogged packet; this user remains blocked until the backlogged packet is transmitted successfully.

Let N^t be a random variable that represents the number of backlogged users at time t. Let S^t represent the channel input rate (i.e., the average rate at which new packets are transmitted over the channel) at time t. Then S^t can be expressed as $S^t = (M - N^t)\sigma$. If M and σ are assumed to be time-invariant, then N^t is a Markov chain with stationary transition probabilities. See [Lam, 1974] for the one-step state transition probabilities.

Consider (N^t, S^t) in the two-dimensional (n, S) plane. Since M and σ are constant, the points (N^t, S^t) must lie on the line $S^t = (M - N^t)\sigma$, which is called the "channel load line." For a fixed K, there exists in the (n, S) plane an "equilibrium contour," defined as the locus of points such that the channel input rate is equal to the expected channel throughput rate. Equilibrium contours for several values of K are shown in Figure 6.1.

Given the above Markovian model, it is possible to describe stable and unstable channels [Lam, 1974]. The channel load line and the equilibrium contour for a given channel may intersect at one or more "equilibrium points." A slotted ALOHA channel is defined to be stable if the intersection of its channel load line and equilibrium contour consists of exactly one (equilibrium) point. Otherwise, the channel is said to be unstable.

In Figure 6.2, a stable channel and an unstable channel are illustrated. The arrows on the channel load lines indicate the directions of flow predicted by the approximation derived in the time-dependent analysis of slotted ALOHA [Lam, 1974]. In the figure, the arrows point toward a decreasing backlog when the throughput rate is greater than the input rate, and an increasing backlog when the throughput rate is less than the input rate. Several equilibrium points are shown; each point that serves as a sink for the drift of N^t is referred to as a stable equilibrium point, while a point that acts as a source is called an unstable equilibrium point.

Shown in Figure 6.2(a) is the channel load line for a stable channel. The sole equilibrium point (n_0, S_0) is stable and is called the channel operating point; N^t will tend to drift toward that point. Note that if the number of users M is finite, it is always possible to construct a stable channel by making the number of retransmission slots K sufficiently large (Figure 6.1.).

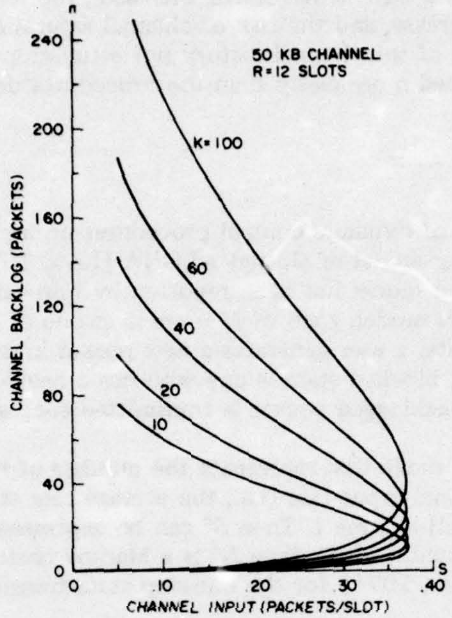


Fig. 6.1—Slotted ALOHA: equilibrium contours on the (n, S) plane [Lam, 1974, p. 103]

Figure 6.2(b) shows the channel load line for an unstable channel. In addition to the desired operating point, there is an additional stable equilibrium point, the channel saturation point, with a very large backlog and virtually zero throughput. A third point at $n = n_c$ is an unstable equilibrium point since the flow is away from it. An unstable channel behaves as follows. Starting with an empty system, the channel performance initially is in the neighborhood of the channel operating point (n_0, S_0). After a finite

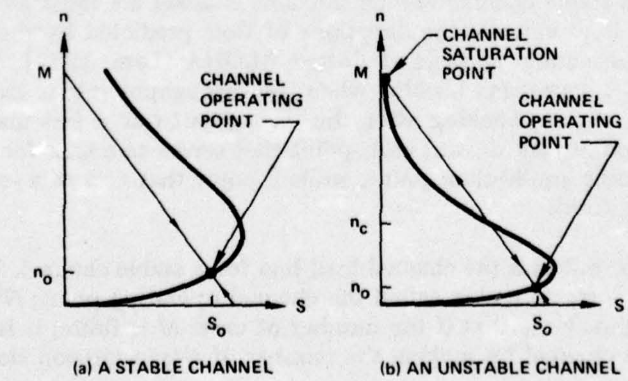


Fig. 6.2—A stable and an unstable slotted ALOHA channel [Lam, 1974, p. 108]

time period, the channel backlog N^t crosses the unstable equilibrium point at $n = n_c$. Once this occurs, N^t moves toward the saturation point, where the backlog increases significantly and the throughput vanishes. When N^t reaches the saturation point, the channel has failed.

Given an unstable channel, there are various ways to convert it into a stable channel; one is to decrease M, the number of users, while another is to increase K, the number of retransmission slots. However, each of these alternatives has serious disadvantages; while the former restricts channel utilization significantly, the latter may increase packet delay to an unacceptable level. A third solution is the use of dynamic control procedures such as those described above.

A brief description of how the ICP, RCP, and IRCP would be applied to the Markovian model of the channel presented above follows. These examples use a "control limit" policy that works in this manner: Assume a number of states 0, 1, 2, . . . , M and two actions a_1 and a_2. Let $\hat{n}$ represent the control limit. Then, if s is the current state,

$$\text{for } 0 \leqslant s \leqslant \hat{n}, \qquad \text{take } a_1$$

$$\text{for } \hat{n} + 1 \leqslant s \leqslant M, \qquad \text{take } a_2 .$$

For these examples, the state is equivalent to N^t, the number of backlogged users at time t.

In Figure 6.3(a), the application of the control limit policy is illustrated for ICP. If $N^t \leqslant \hat{n}$, then $S^t = (M - N^t)\sigma$; if $N^t > \hat{n}$, then $S^t = 0$. Figure 6.3(b) shows the application of this policy for RCP. If $N^t \leqslant \hat{n}$, then $K = K_0$, and if $N^t > \hat{n}$, then $K = K_c$. For IRCP, two control limits $\hat{n}_1$ and $\hat{n}_2$ exist with $\hat{n}_1 < \hat{n}_2$. IRCP works as follows. For $0 \leqslant N^t \leqslant \hat{n}_1$, all new packet transmission requests are accepted and a retransmission interval $K = K_0$ is used. For $\hat{n}_1 < N^t \leqslant \hat{n}_2$, all new packet transmission requests are accepted, but $K = K_c$. For $N^t > \hat{n}_2$, all new packet transmission requests are rejected and K is set equal to K_c.

Lam has formulated a second model that allows the cost of control procedures such as ICP and RCP to be determined [Lam, 1974]. He shows that an optimal control policy

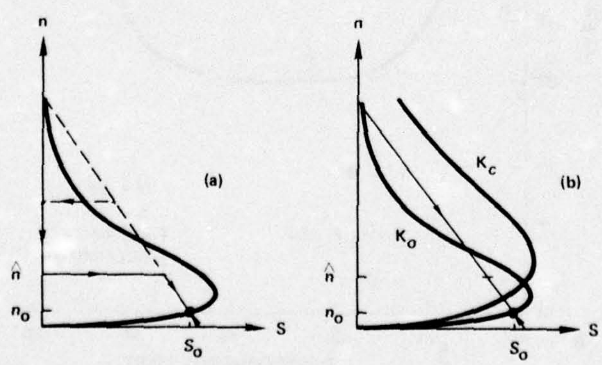


Fig. 6.3—Control limit policy examples [Lam 1974, p. 149]; (a) ICP and (b) RCP

exists which maximizes the channel throughput and minimizes the average delay simultaneously. Lam also presents an efficient algorithm called POLITE that, given a channel load line and a dynamic control procedure, finds the optimal control policy and computes the optimum channel performance measures.

The throughput-delay performance of a slotted ALOHA channel operated under the ICP and RCP control limit policies has been analyzed by Lam [Lam, 1974]. This analysis assumes a satellite channel with a propagation delay of 12 slots and that users have perfect knowledge of the number of backlogged users. The operating value of K is 10 slots, since this value gives a channel operating point close to the optimum. Each channel load line used in the computations is specified by M, the number of users, and an operating point (n_0, S_0). The operating points used in Figures 6.4-6.8 are (4, 0.32) and (7, 0.36).

In Figures 6.4 and 6.5, the channel performance of the ICP and RCP control limit policies for M = 200, S_0 = 0.32 and 0.36 is shown. For both ICP and RCP, a single control limit maximizes throughput and minimizes delay. Note the flatness of the curves, especially for S_0 = 0.32, near the optimum performance point. This flatness implies that

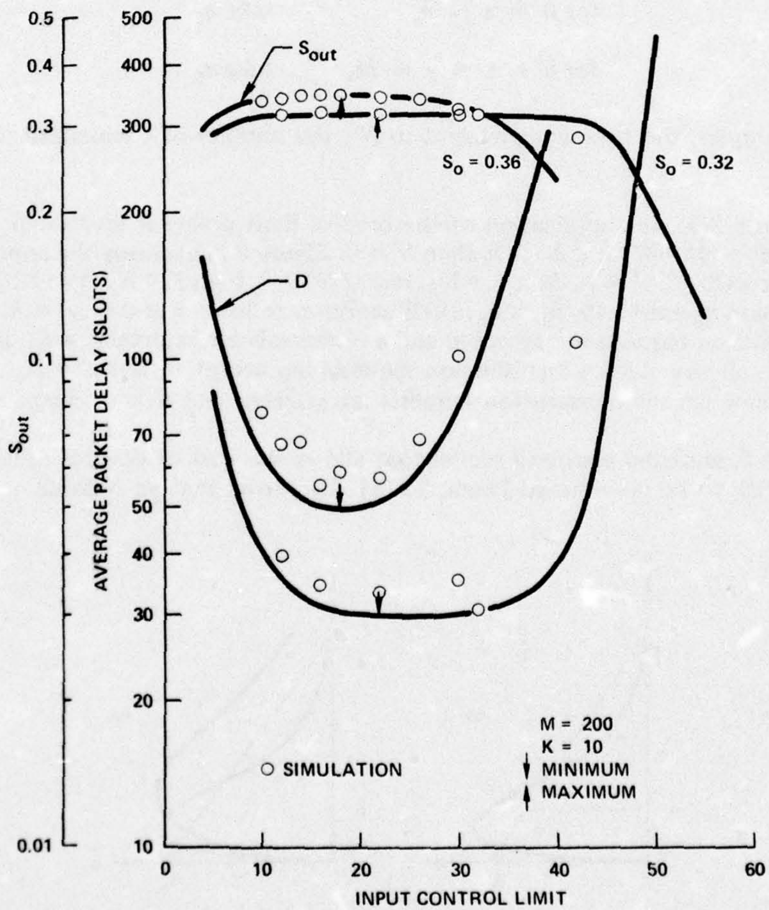


Fig. 6.4—Slotted ALOHA: channel performance vs ICP control limit for M = 200 [Lam, 1974, p. 188]

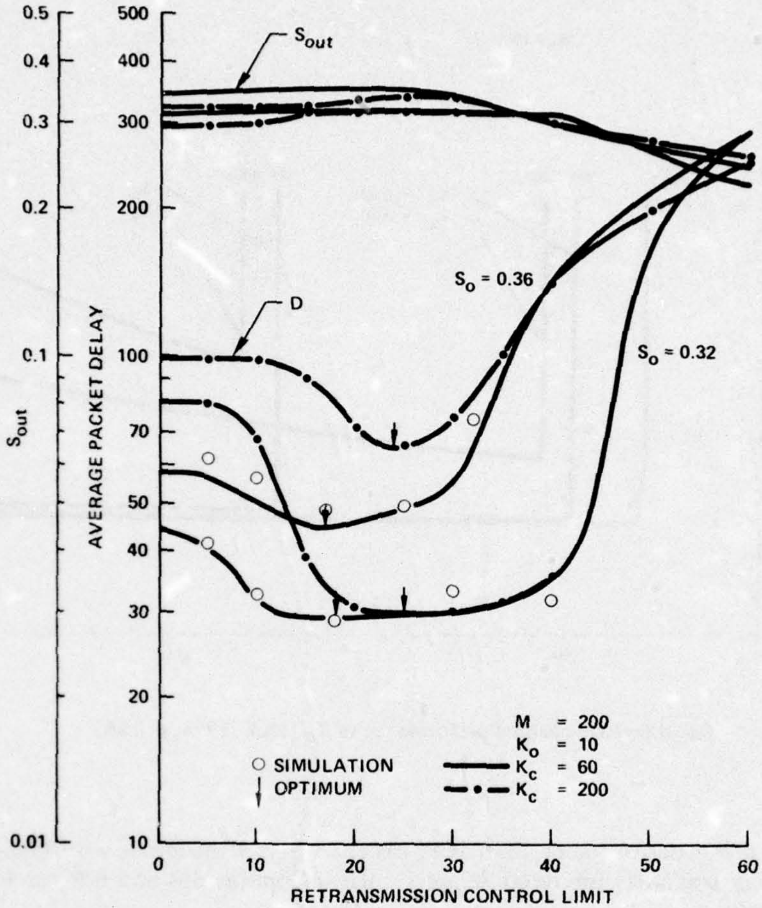


Fig. 6.5—Slotted ALOHA: Channel performance vs RCP control limit for M = 200 [Lam, 1974, p. 189]

some margin for error exists; i.e., if, as in practical systems, the users' knowledge of the number of backlogged users is imperfect, it is still possible to achieve near optimum throughput-delay performance.

In a practical system using RCP, the values of M and σ will vary in time. Thus the value of K_c, the number of retransmission slots used when the control limit is exceeded, must handle a range of user input rates and/or population sizes. Figure 6.6 shows how the value of K_c affects delay performance. Note the disastrous effect on delay that occurs when K_c is too small. Making K_c overly large, however, may cause a significant increase in average packet delay. Figure 6.5 shows the degradation in channel performance that takes place when K_c = 200 as opposed to K_c = 60.

In Figure 6.7(a), the optimal control limit for ICP and RCP is shown for different values of M. Increasing the number of users M has minimal effect on the control limit for ICP. Figure 6.7(b) shows the effect of a change in M on packet delay. RCP provides

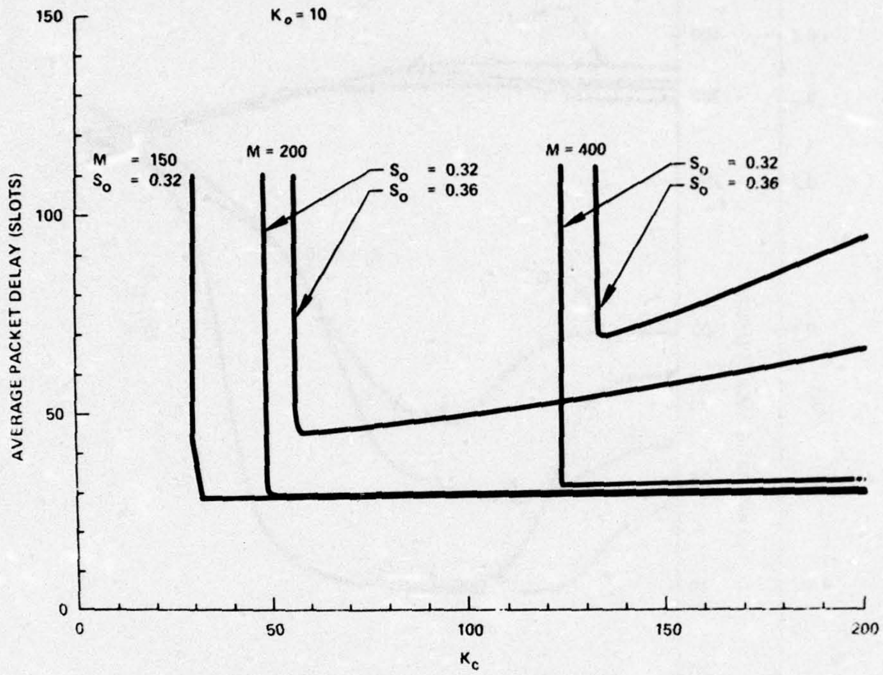


Fig. 6.6—RCP channel performance vs K_c [Lam, 1974, p. 186]

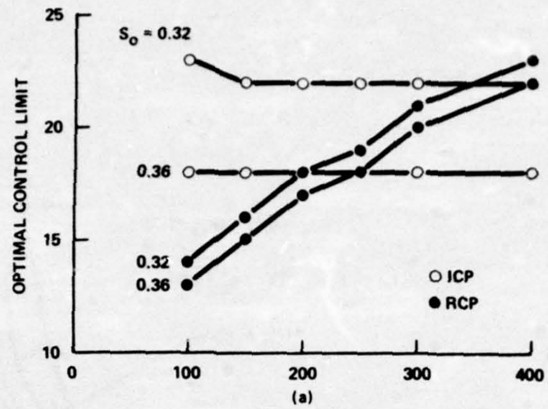
slightly better delay performance than ICP, except when M becomes very large. Optimum throughput-delay tradeoffs for fixed M are shown in Figures 6.8 and 6.9 for ICP and RCP, respectively.

The optimum control limit for ICP is 22. For both ICP and RCP, the optimum channel performance is very close to that of the infinite population model. In fact, the optimum throughput-delay performance for the case $M = 50$ is superior to that of the infinite population model. This is because a user population of 50 gives rise to a stable channel and hence performance at the operating point is achieved.

Since IRCP enforces control policies for both ICP and RCP, its performance should be as good as or better than that given by ICP and RCP. In fact, in the cases that have been analyzed, IRCP consistently gives the best performance. See Table 6.1 for a comparison of the performance of the three procedures.

6.3 Simulation

Lam [Lam, 1974] has reported simulation results (a) for ICP and RCP using optimal control policies and assuming perfect channel state information, (b) for the CONTEST algorithm using optimal ICP and RCP policies, and (c) for heuristic RCP.



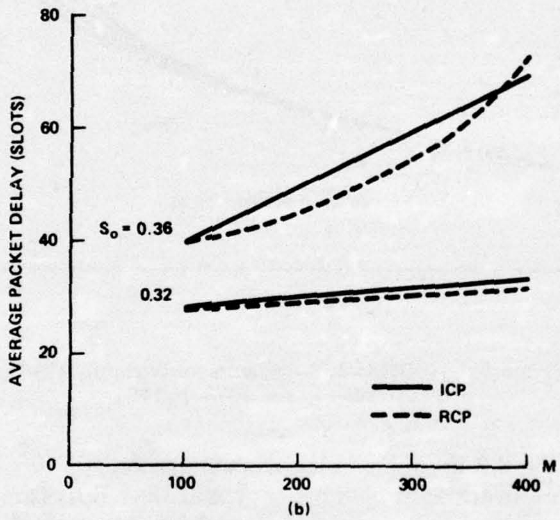


Fig. 6.7—Slotted ALOHA: ICP and RCP channel performance vs M [Lam, 1974, p. 193]

The agreement between the simulation results and the analytic results described in the foregoing section is good. See Figures 6.4 and 6.5 for a comparison of the analytic and simulation results for ICP and RCP. The simulation results show further that application of either the CONTEST algorithms or heuristic RCP does not lead to serious deterioration in channel performance. Although both sets of control algorithms give results close to that predicted by POLITE, the CONTEST algorithms tend to produce slightly better performance than heuristic RCP.

Lam has shown through simulation [Lam, 1974] that an input rate of 0.8 packet/slot sustained for 100 time slots is sufficient to saturate an uncontrolled slotted ALOHA

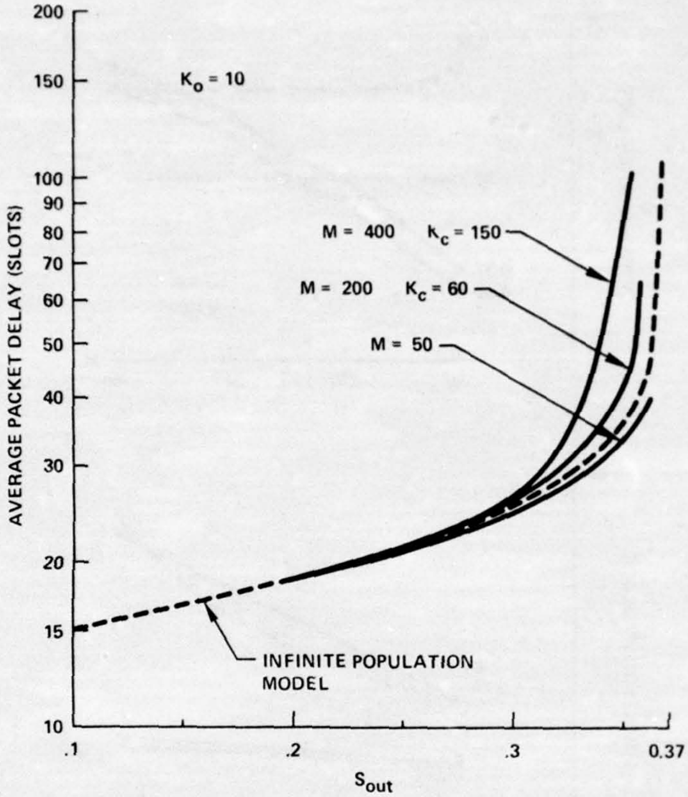


Fig. 6.8—Slotted ALOHA: ICP optimum throughput-delay trade-offs at fixed M [Lam, 1974, p. 196]

channel. Shown in Table 6.2 are the results of a simulation run in which heuristic RCP was applied to a heavily overloaded channel. For 200 time slots (the time period 1001-1200 in the table), an input rate of one packet per slot was applied to the channel. The algorithm was able to handle the serious traffic overload that resulted. Whereas, in an uncontrolled channel, the throughput rate would vanish under such heavy traffic; under this control procedure, a throughput rate of around 30% was maintained in spite of the overload. Moreover, within 3000 time slots, channel operation had returned to nearly normal. Simulation studies have shown similar behavior for IRCP.

6.4 Discussion

To be able to guarantee an acceptable level of system performance in a random access system, some form of dynamic channel control is required. The above discussion shows that in addition to preventing channel saturation, the application of these policies allows channel performance very close to the theoretical optimum. It should be emphasized, however, that these procedures are designed to handle temporary rather than long-term increases in the channel input. Since, in practical systems, the channel input may vary considerably over long time periods, the use of additional control mechanisms may be necessary.

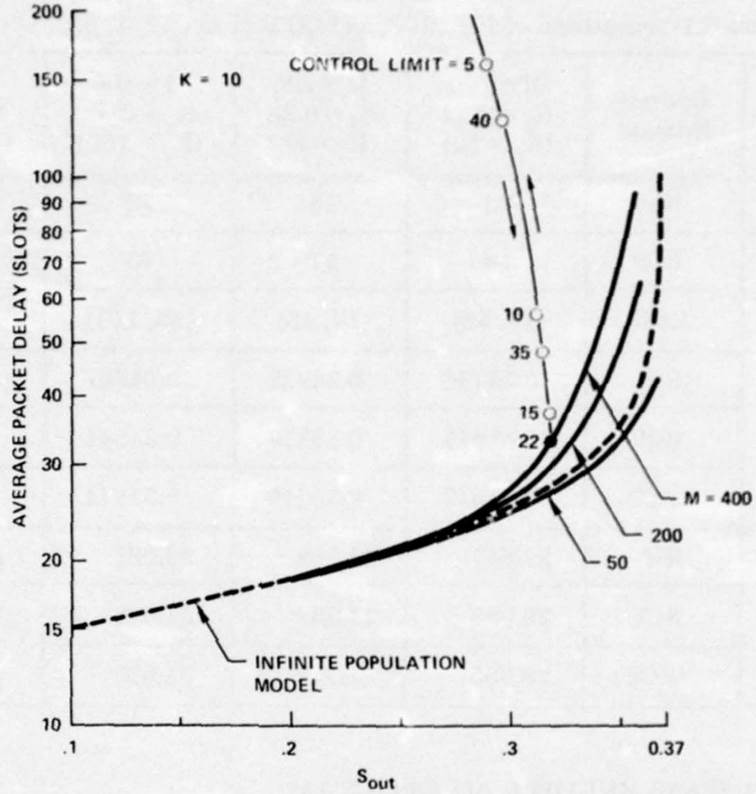


Fig. 6.9—Slotted ALOHA: RCP optimum throughput-delay tradeoffs at fixed *M* [Lam, 1974, p. 197]

Of concern in the choice of control schemes is the difficulty of implementation. Heuristic RCP is clearly more easily implemented than the ICP and RCP policies, which require estimation of the channel state.

Recall that for the infinite population model, the theoretical maximum throughput rate which can be achieved on a slotted ALOHA channel is still only 36%. It must be emphasized that dynamic control procedures do not increase the throughput rate achievable on the channel from its theoretical maximum. These procedures reduce significantly the risk of channel saturation.

The slotted ALOHA scheme is not the only scheme with which dynamic control procedures may be used. These procedures are equally effective with other contention schemes such as pure ALOHA and Carrier Sense Multiple-Access which, like slotted ALOHA, are vulnerable to unstable behavior.

Table 6.1 Comparison of ICP, RCP, and IRCP [Lam, 1974, p. 201]

Parameter	Control Scheme	M = 200 $S_0 = 0.32$ ($K_c = 60$)	M = 200 $S_0 = 0.36$ ($K_c = 60$)	M = 400 $S_0 = 0.32$ ($K_c = 150$)	M = 400 $S_0 = 0.36$ ($K_c = 150$)
$\hat{n}$	ICP	22	18	22	18
$\hat{n}$	RCP	18	17	23	22
$(\hat{n}_1, \hat{n}_2)$	IRCP	(18, 56)	(17, 43)	(23, 116)	(23, 91)
S_{out}	ICP	0.31778	0.34925	0.31807	0.34846
	RCP	0.31817	0.35217	0.31844	0.34715
	IRCP	0.31817	0.35219	0.31844	0.34847
D	ICP	29.857	49.552	33.096	69.237
	RCP	29.085	44.802	31.608	73.588
	IRCP	29.085	44.772	31.608	69.215

7.0 CARRIER SENSE MULTIPLE ACCESS (CSMA)

7.1 Description

In ALOHA systems, the major factor limiting channel capacity is the collision of packets from different users. By decreasing the probability of a packet collision, a higher channel capacity may be obtained. When the propagation delay between any source-destination pair is small, such as in ground radio situations, the Carrier-Sense Multiple-Access (CSMA) technique can be used to help avoid collisions. This approach was proposed and described by Tobagi and Kleinrock [Tobagi, 1974; Kleinrock & Tobagi, 1975a, b]. In CSMA, a user attempts to avoid collisions by listening for the presence of the carrier due to another user's transmission, and then takes action based on the sensed state of the channel.

Three different protocols, 1-persistent, non-persistent, and p-persistent CSMA have been considered. They differ in the action that a user takes after sensing the channel. In all cases, when a user discovers that a transmission has been unsuccessful, it reschedules the transmission according to a delay selected from a retransmission delay distribution. At this randomly selected time, the user reinstitutes the specific protocol being used.

Table 6.2—Simulation Run for Heuristic RCP Subject to Channel Input Pulse*†
[Lam & Kleinrock, 1975, p. 903]

Time Period	Throughput Rate	Traffic Rate	Average Delay	Average Backlog
1 - 200	0.285	0.395	19.8	2.1
201 - 400	0.320	0.390	16.3	1.2
401 - 600	0.255	0.425	22.8	2.8
601 - 800	0.290	0.475	26.1	4.0
801 - 1000	0.325	0.570	28.5	5.7
1001 - 1200	0.230	2.395	34.1	68.8
1201 - 1400	0.285	1.695	141.3	112.6
1401 - 1600	0.310	1.500	273.1	91.8
1601 - 1800	0.375	1.415	288.6	68.5
1801 - 2000	0.280	1.110	224.6	53.1
2001 - 2200	0.360	1.240	257.3	48.8
2201 - 2400	0.355	0.925	193.9	31.3
2401 - 2600	0.385	0.655	122.8	15.2
2601 - 2800	0.320	0.565	68.0	8.8
2801 - 3000	0.280	0.420	39.3	5.6
3001 - 3200	0.295	0.495	31.6	6.3
3201 - 3400	0.265	0.680	45.0	11.7
3401 - 3600	0.350	0.750	37.0	13.3
3601 - 3800	0.310	0.465	65.2	8.2
3801 - 4000	0.275	0.520	33.6	7.7
4001 - 4200	0.330	0.480	34.6	5.2
4201 - 4400	0.325	0.615	29.5	7.5
4401 - 4600	0.370	0.525	38.6	7.6
4601 - 4800	0.260	0.705	44.2	15.9
4801 - 5000	0.375	0.720	63.5	11.1
5001 - 5200	0.350	0.635	41.7	9.0
5201 - 5400	0.285	0.475	29.3	6.6
5401 - 5600	0.315	0.510	30.4	4.9
5601 - 5800	0.290	0.425	24.1	4.1
5801 - 6000	0.305	0.490	28.7	4.7

*Average values in 200 time slot periods.
†Input parameters:
Number of terminals M = 400, propagation delay R = 12
For the time period 1 - 1000, input rate $M\sigma = 0.3232$
For the time period 1001 - 1200, input rate $M\sigma = 1.0$
For the time period 1201 - 6000, input rate $M\sigma = 0.3232$
$K_1 = 10$ $K_n = 150$ $(m \geqslant 2)$

7.1.1 1-Persistent CSMA

This protocol is designed to achieve greater throughput by never letting the channel go idle if there is a user terminal with a packet ready to transmit (a ready terminal). In this technique:

- If the channel is sensed idle, the terminal transmits the packet.
- If the channel is sensed busy, the terminal continues sensing the channel until the channel goes idle and then immediately transmits the packet.

A slotted version of this protocol can be considered in which time is slotted with slot size τ = propagation delay. All terminals are synchronized and transmissions are constrained to begin only at the beginning of a slot. If a terminal becomes ready during some slot, it senses the channel at the beginning of the next slot and then operates under the above protocol.

7.1.2 Non-Persistent CSMA

The previous protocol tends to minimize channel idle time. However, if two or more terminals find the channel busy, they all wait until the channel is idle, transmit, and suffer a collision with probability 1.0. Non-persistent CSMA limits this interference but may introduce idle periods. In this protocol

- If the channel is sensed idle, the ready terminal transmits the packet.
- If the channel is sensed busy, the ready terminal schedules the transmission after a random retransmission delay. At this new time, the terminal senses the channel and repeats the protocol.

A slotted version of this protocol is also possible.

7.1.3 p-Persistent CSMA

The two previous protocols differ according to whether or not attempted transmissions are rescheduled when the channel is sensed busy. The first protocol does not reschedule, with probability one; the second does not reschedule, with probability zero. The former reduces idle time with an increase in the chance of collision; the latter decreases the chance of collision but can increase the channel's idle time. The third protocol, p-persistent CSMA, has been proposed in order to take advantage of the good properties of each of the other protocols. This protocol uses a randomization parameter p, where $0 \leqslant p \leqslant 1$. Time is slotted, with the width of a slot being τ, the maximum propagation time. In this technique

- If the channel is sensed idle, then

    (a) with probability p, the ready terminal transmits.

(b) with probability 1-p, the ready terminal delays the transmission for one slot. If at this new point in time the channel is sensed idle, the same process is repeated. If the channel is busy at this point, the terminal reschedules transmission in accordance with the retransmission delay distribution.

- If the channel is initially sensed busy, then the terminal waits until the channel is sensed idle and at that point operates as above.

For $p = 1$, this protocol is the same as 1-persistent CSMA.

7.2 Analysis

The results presented here can be found in [Tobagi, 1974; Kleinrock and Tobagi, 1975a, b]. In the analysis, it is assumed that unsuccessful packet receipt by a station is caused by a packet collision and not by noise on the channel. In addition, all terminals are assumed to be within line-of-sight range of one another. The situation where terminals are hidden from the central station, necessitating repeaters and network considerations, has not been considered. It is also assumed that a terminal may be receiving or transmitting, but not both simultaneously; turnaround time, however, is considered to be negligible.

The traffic source is considered to consist of a very large number of users who collectively form an independent Poisson source with a mean packet generation rate of λ packets/second. Each user delays retransmission of a previously collided packet by some random time whose mean $\overline{X}$ is assumed to be large compared to T, the packet transmission time (packets are assumed to be the same length). Furthermore, it is assumed that the process defined by the start times of new packets as well as previously collided packets is a stationary independent Poisson process. In addition, each user is assumed to have at most one packet requiring transmission at any given time.

The analysis, as with pure and slotted ALOHA, considers the relationship between S, the average channel throughput, and G, the average channel traffic (new packets plus collisions). Basic equations for S are derived in terms of G and $a = \tau/T$, where τ is the maximum source-destination propagation time and T is the packet transmission time. The equations can be found in the references cited. The maximum throughput for an access mode is defined to be the capacity of the channel under the specified mode. Table 7.1 summarizes the channel capacity for the various protocols considered ($a = 0.01$ is used). Pure and slotted ALOHA are included for comparison.

In Figure 7.1, S vs. G is plotted for all the above protocols.

While the capacity of a pure ALOHA or slotted ALOHA channel is independent of the propagation delay, the CSMA channel capacity is dependent on $a = \tau/T$. In Figure 7.2, channel capacity for the various protocols is plotted as a function of a. For a value of a close to 1.0, slotted and even pure ALOHA are superior to CSMA, because decisions based on partially obsolete data are deleterious.

Figure 7.3 is a comparison of the various access modes in terms of the average number of transmissions required per packet, which is proportional to the average delay.

Table 7.1—Channel Capacity for Various Protocols [Kleinrock and Tobagi, 1975b, p. 1433]

Protocol	Capacity C
Pure ALOHA	0.184
Slotted ALOHA	0.368
1-Persistent CSMA	0.529
Slotted 1-Persistent CSMA	0.531
0.1-Persistent CSMA	0.791
Non-Persistent CSMA	0.815
0.03-Persistent CSMA	0.827
Slotted Non-Persistent CSMA	0.857
Perfect Scheduling	1.000

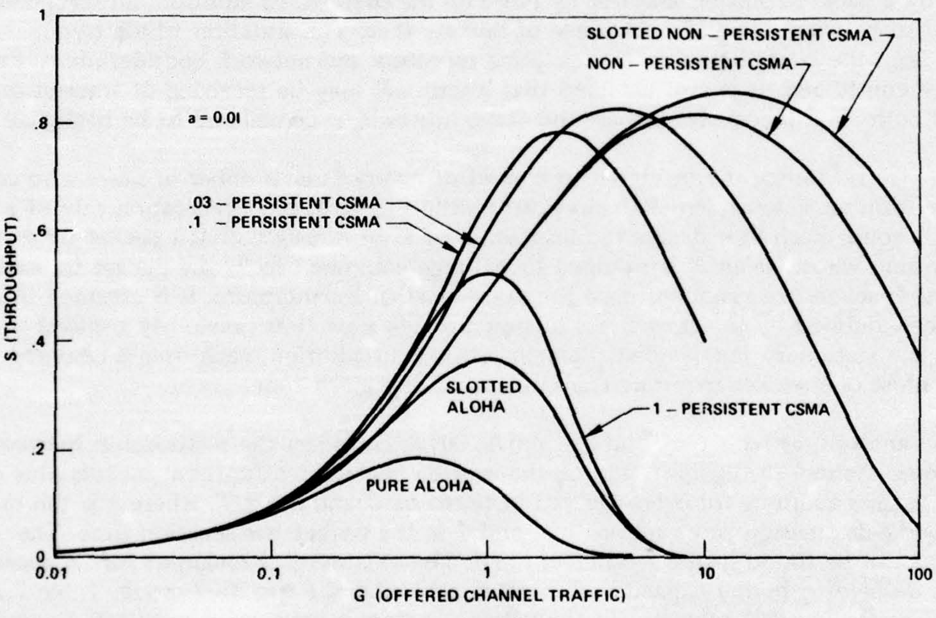


Fig. 7.1—Channel traffic vs channel throughput for the various ALOHA modes (a = 0.01) [Tobagi, 1974, p. 75]

Again, a is taken to be 0.01. Note that the CSMA modes provide a lower average number of transmissions per packet than the two ALOHA modes. Also, for each value of S, there is an optimum value of p such that p-persistent is best. For small values of S, $p = 1$ is optimum (i.e., 1-persistent CSMA). As S increases, the optimal value for p decreases.

7.3 Simulation

The average delay D is a function of both the channel throughput S and the mean retransmission delay $\bar{X}$. For each value of S, a minimum delay can be achieved by

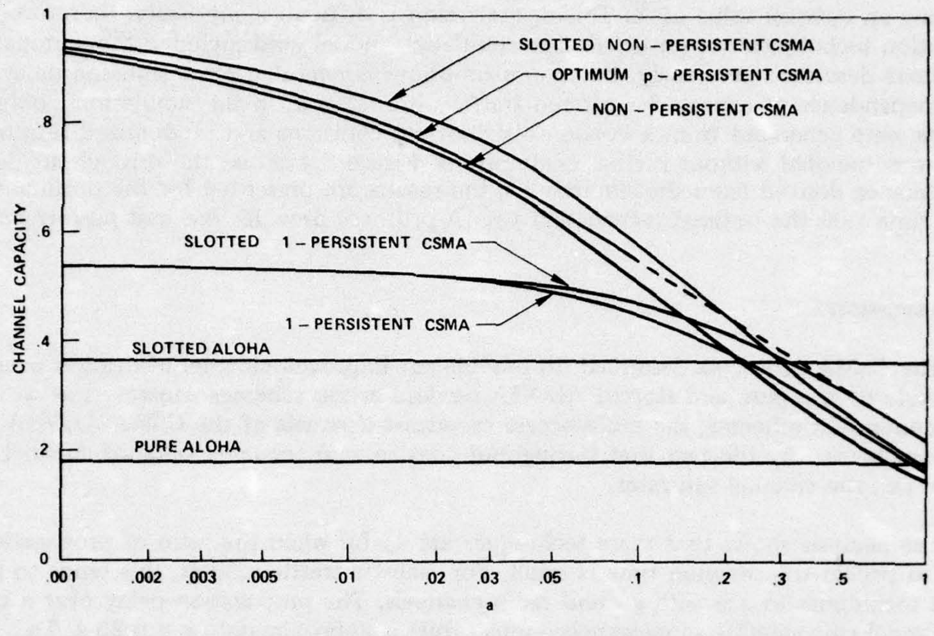


Fig. 7.2—**CSMA** *vs slotted* **and pure ALOHA: effect of propagation delay on channel capacity [Tobagi, 1974, p. 77]**

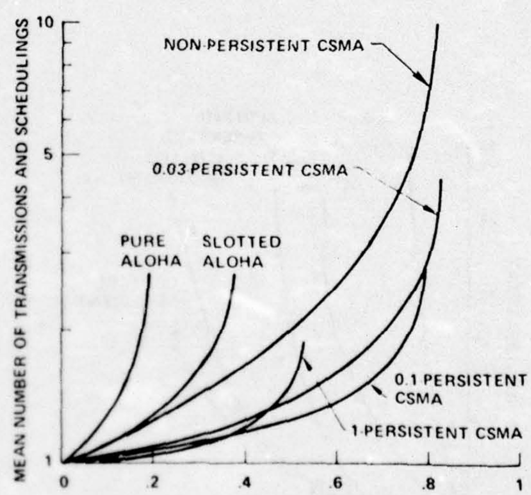


Fig. 7.3—CSMA vs pure and slotted ALOHA: G/S vs throughput (a = 0.01) [Tobagi, 1974, p. 81]

choosing an optimal value of $\overline{X}$. This optimization is difficult analytically; therefore, simulation techniques were utilized. The simulation model used included the various assumptions described previously, but the assumptions concerning retransmission delay and the independence of arrivals for offered traffic were relaxed. In the simulations, only new packets were generated from a Poisson distribution; collisions and randomized retransmissions were handled without further assumptions. Figure 7.4 shows the throughput/delay performance derived from the simulations; the results are presented for the optimal values of $\overline{X}$. Note that the optimal p-Persistent CSMA protocol provides the best performance.

7.4 Discussion

The CSMA protocols described do provide for improved channel utilization over that achievable by the pure and slotted ALOHA random access schemes. However, as with most contention schemes, the multi-access broadcast channels of the CSMA-ALOHA type are characterized by the fact that throughput goes to zero for large value of channel traffic; i.e., the channel saturates.

The analysis shows that these techniques are useful when the ratio of propagation delay to packet transmission time is small. For realistic traffic models, this tends to limit CSMA techniques to use with ground radio channels. The propagation delay over a satellite channel (the satellite in geosynchronous orbit) is approximately τ = 0.25 s. To achieve a small value of $a = \tau/T$ leads to considering long packet transmission times and long vulnerable periods for collision. Thus, this set of protocols is most likely to be applicable in the ground radio environment for which it was originally proposed.

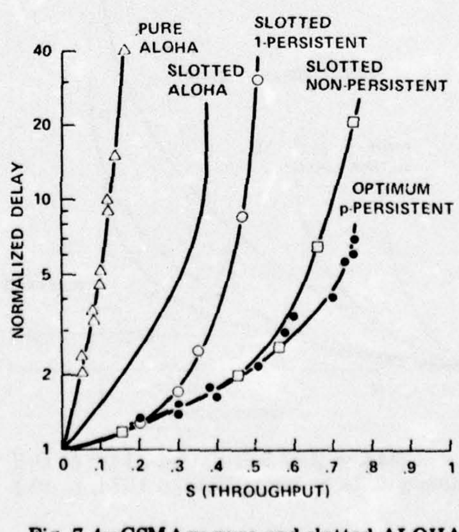


Fig. 7.4—CSMA vs pure and slotted ALOHA: throughput-delay tradeoffs from simulation (a = 0.01) [Tobagi, 1974, p. 83]

The analysis of CSMA depends upon the assumption that all terminals are in radio line-of-sight of each other. Tobagi and Kleinrock have also examined the "hidden terminal" case, where terminals are in range of a central station but are not necessarily in line-of-sight of each other [Tobagi, 1974; Tobagi & Kleinrock, 1975]. Their analysis shows that the presence of hidden terminals in the CSMA scheme can significantly degrade the achievable channel capacity.

As a solution to this problem, Tobagi and Kleinrock have proposed a protocol called Busy Tone Multiple-Access (BTMA), which is based on the assumption that all terminals are within line-of-sight of the central station, if not of each other. A small communications channel is designed as the "busy tone" channel. The central station transmits a busy tone signal on the small channel when it senses the presence of the carrier on another (large) channel used by all terminals to transmit packets. The terminals monitor the busy tone channel and, based on the presence or absence of a signal on that channel, operate as in CSMA. The performance of BTMA is very similar to that of the CSMA protocols. For details of the analysis, see Tobagi [Tobagi, 1974; Tobagi & Kleinrock, 1975].

8.0 ROBERTS' RESERVATION

8.1 Description

In order to accommodate data traffic composed of multi-packet as well as single-packet messages, Roberts [Roberts, 1973] has proposed a technique in which the channel is operated in slotted ALOHA mode part of the time, when stations make reservations, and in dedicated mode for the rest of the time, when stations send messages composed of one or more packets. As in slotted ALOHA, the channel is divided into time slots of fixed-length T, where T is equal to the duration of a packet transmission. The channel is operated in two states, the "reservation state" and the "ALOHA state." When the channel is in the reservation state, a frame consisting of $M + 1$ slots is used. While the last M slots of the frame are used for the transmission of message packets, the first slot in the frame is subdivided into V "small slots." These small slots are accessed by all stations on a contention basis, where the form of contention used is slotted ALOHA random access. Each station uses these small slots to reserve the number of slots needed to transmit a given message and to receive acknowledgments of correct message transmission. Whenever a reservation is made, all stations add the number of slots requested to a count of the number of slots currently reserved. In this way, each station always knows when the first unreserved slot will occur so that, when it reserves slots for its own use, it knows when to begin transmitting without having to keep track of every previous reservation. Thus, for this protocol, there exists a single, distributed queue which contains requests for use of the non-ALOHA (dedicated) slots. Each station can get its message packets into the queue by broadcasting a reservation. Note that the queue length can exceed M, a value that determines only the interval between the times at which stations can add reservations to the queue.

Whenever the queue length becomes zero, the channel switches from reservation state to ALOHA state. In the ALOHA state, every slot (instead of every $(M + 1)$st slot) is subdivided into V small slots that are available for making reservations using slotted ALOHA. When the next valid reservation is made, the channel reverts to the reservation

state and operates as described in the previous paragraph. For an example of how Roberts' reservation scheme works, see Figure 8.1.

8.2 Analysis

Roberts' analysis of this scheme is based on the following traffic model. Messages are composed of either a single packet or eight packets, where a packet is of fixed length Each station generates messages with a Poisson arrival rate. A fraction F of these messages is assumed to consist of a single packet, while the remaining messages are assumed to contain eight packets each. For the analysis, a value $F = 0.5$ is used.

Roberts has compared the throughput-delay performance of this reservation scheme to that of alternative schemes. For the reservation scheme, the average delay associated with an average size message is computed, and this delay is compared with the analogous delays for slotted ALOHA and TDMA (note that the performance measure used is *message* delay rather than packet delay). For channel utilizations below 0.15, the average delay for slotted ALOHA is less than that of the reservation scheme; for channel utilizations above 0.15, the reservation technique has less average delay than slotted ALOHA. Moreover, Roberts concludes that for any channel utilization, the average message delay of the reservation scheme is lower than that associated with TDMA. He has also investigated the relation of cost to delay, the effect of station traffic on cost, and the effect of the number of stations on cost for the reservation scheme, slotted ALOHA, TDMA, FDM, and FDM with "store-and-forward star." In each case, the reservation scheme is shown as least expensive. However, since the foregoing results are accompanied by few details, the conclusions given are somewhat difficult to evaluate.

8.3 Simulation

None.

8.4 Discussion

When the user population consists of many small stations whose messages are predominately single-packet messages, both pure ALOHA and slotted ALOHA can provide good performance. However, when the traffic contains a significant portion of multipacket messages, a reservation scheme such as Roberts' is more appropriate. As indicated

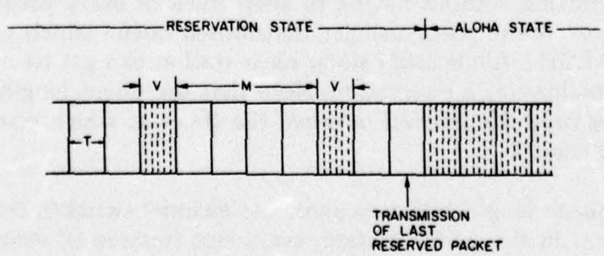


Fig. 8.1—Roberts' Reservation technique: channel structure

above, for a channel utilization above 0.15, the delay performance of the reservation scheme is superior to that of slotted ALOHA. Moreover, while the maximum channel utilization for the slotted ALOHA scheme is only 36%, a much higher channel throughput rate can be sustained by the reservation technique.

Regarding stability, the same phenomena that cause a pure or slotted ALOHA channel to be unstable also have this effect on a reservation ALOHA channel. In the case of reservation ALOHA, however, it is a high message rate (rather than a high total channel traffic rate) that will cause unstable behavior. Here again, reservation ALOHA should be advantageous, provided that the average message is long compared to its reservation time, and that traffic fluctuations are reflected more as changes in message lengths than as changes in the rate at which messages are generated.

9.0 SPLIT-CHANNEL RESERVATION MULTIPLE ACCESS (SRMA)

9.1 Description

The Split-Channel Reservation Multiple Access (SRMA) technique has been proposed by F. A. Tobagi [Tobagi, 1974]. Under SRMA, the channel is split into two smaller channels; one of these, the "control" channel, is used for the transmission of control information, while the other, the "message" channel, is used for the transmission of the messages themselves. The split into two channels may be achieved using either a time-division or frequency-division scheme.

Two versions of SRMA have been suggested: the Request-Answer to Request-Message technique (R.A.M.) and the Request-Message technique (R.M.). Both versions assume the existence of a scheduling station which receives on the control channel "requests-to-transmit" from other stations and then schedules messages for transmission on the message channel.

9.1.1 Request-Answer to Request-Message (R.A.M.) Technique

Under the R.A.M. scheme, the control channel is further divided into two channels: the "request" channel and the "answer-to-request" channel. A station with a message to transmit accesses the request channel in a random access mode; pure ALOHA, slotted ALOHA, and Carrier-Sense Multiple-Access (CSMA) are three alternative ways of operating the request channel. The request is made via a request packet which includes the identification of the requesting station and, in the case of variable-length or multi-packet messages, the message length. Upon receiving a request packet, the scheduling station uses the answer-to-request channel to transmit an "answer" packet, i.e., a packet which contains the requesting station's ID along with the time at which that station can initiate transmission of its message. At the assigned time, the requesting station sends its message via the message channel.

Since the request channel is operated in a contention mode, the transmission of a request packet may not be successful. To handle this situation, a requesting station, upon transmission of a request to transmit, executes a time-out. If no answer packet is received by the end of the time-out, it may be assumed that a collision occurred in the request

channel. In accordance with the ALOHA or CSMA protocol, the station, after a randomized retransmission interval, retransmits its request.

9.1.2 Request-Message (R.M.) Technique

Under the R.M. scheme, only two channels, the control channel and the message channel, are required. The control channel is operated exactly like the request channel in R.A.M. The scheduling station, upon receiving the request, queues it. When the message channel is available, the scheduling station uses the message channel to transmit an answer packet containing the requesting station's ID. Upon hearing its ID, the requesting station initiates transmission of its message on the message channel.

9.2 Analysis

The throughput and delay performance of the R.A.M. version of SRMA has been analyzed by Tobagi. He has also completed an approximate analysis of R.M. which indicates that its performance is very similar to that of R.A.M. To determine the performance of R.A.M., the delay D incurred by a message is broken into two components (Figure 9.1):

(a) D_1, the time needed for the receipt of the request packet by the scheduling station, and

(b) D_2, the time between the receipt of the request packet and the end of message transmission.

Since the request channel is operated in random access mode, (i.e., pure ALOHA, slotted ALOHA, or CSMA), D_1 can be computed using the results of previous analysis. See Sections 4 and 7 and [Tobagi, 1974] for these results.

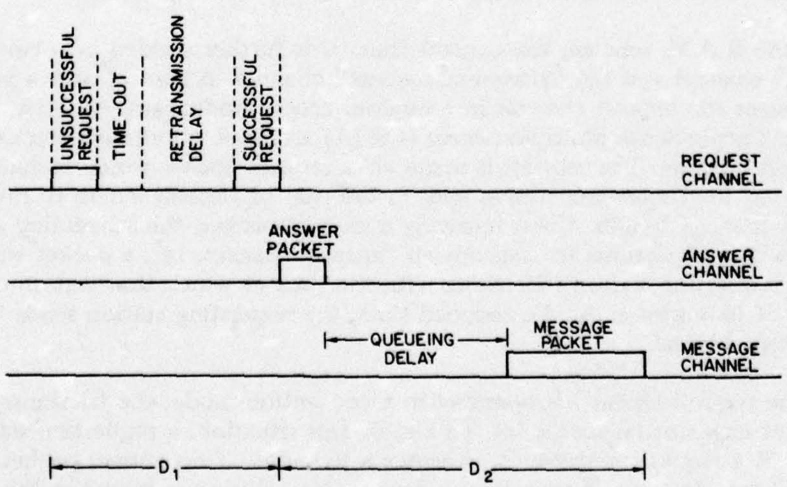


Fig. 9.1—Split-channel reservation Multiple Access [Tobagi, 1974, p. 215]

To estimate D_2, it is assumed that the arrival of successful requests to the scheduling station is Poisson. Then, the message channel can be modeled as a single-server queueing system, in which the arrival process is the Poisson output of the request channel, shifted in time by the transmission time of the answer packet plus propagation delays. Note from Figure 9.1, the answer packet incurs no queueing delays. This is because the scheduling station is the only transmitter on this channel and because it is possible to assign the answer channel enough bandwidth so that an answer packet experiences no queueing delay. In [Tobagi, 1974], D_2 is computed for messages of fixed length and for exponentially distributed message length.

In Figure 9.2, the maximum channel throughput of SRMA and other selected channel management schemes is plotted relative to η, the ratio of the length of a request packet to the length of a message packet. In this figure, as well as in Figure 9.3, a ground radio channel is assumed; moreover, a represents the ratio of propagation delay to average message transmission time, while the expression $\tau W/b_m$ represents the number of messages per time slot, where the length of a time slot is equivalent to the propagation delay. Note that for very small values of η, the theoretical channel capacity of SRMA is close to unity. Operating the request channel in Carrier-Sense (CS) mode rather than in ALOHA mode for $\eta > 0.01$ results in a significant improvement in channel capacity. In comparing the capacity of SRMA to the capacity of the random access modes, SRMA is superior only for relatively small values of η.

In Figure 9.3, the minimum delay for ALOHA-SRMA and Slotted Carrier-Sense SRMA is shown as a function of S, the input rate, for several values of η. Note again the superior preformance that results when the CS mode rather than the ALOHA mode is used to access the request channel; CS does especially well for large values of η. Also shown in Figure 9.3 is the delay performance for CSMA and BTMA. Here, a value of S

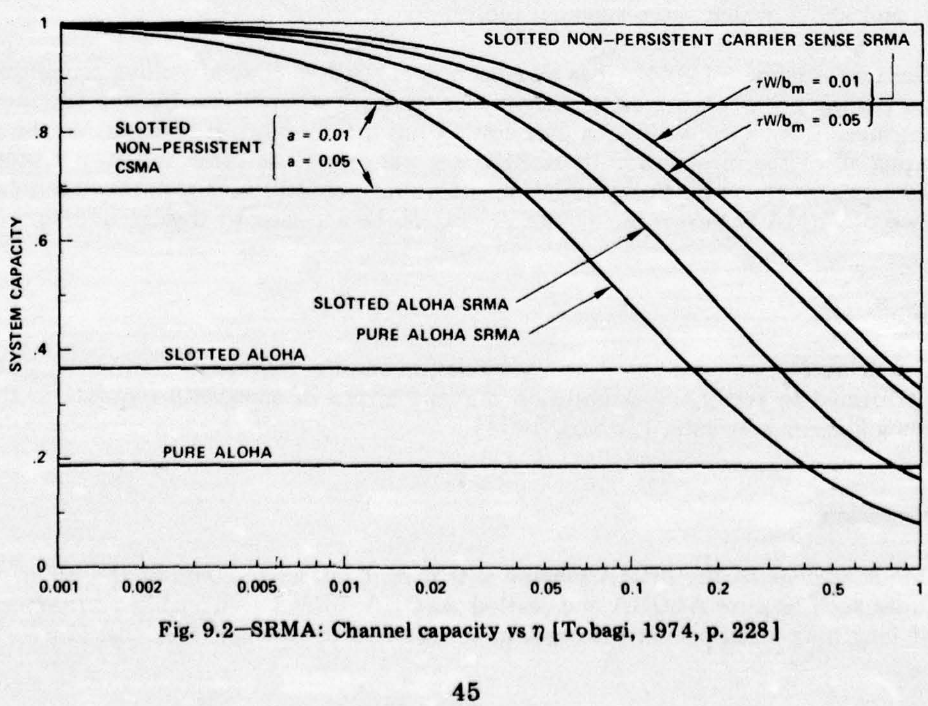


Fig. 9.2—SRMA: Channel capacity vs η [Tobagi, 1974, p. 228]

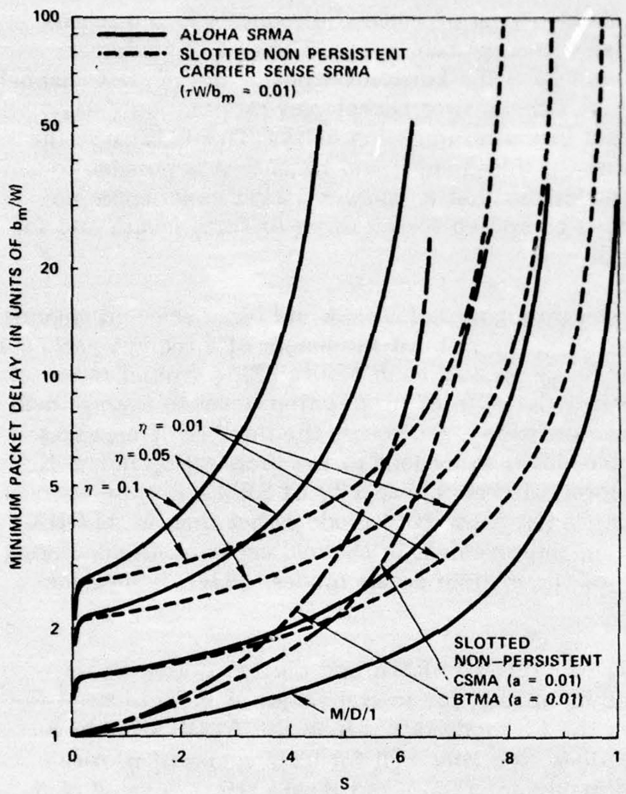


Fig. 9.3—Minimum packet delay in SRMA [Tobagi, 1974, p. 231]

exists below which CSMA or Busy Tone Multiple-Access (BTMA) performs better than SRMA, and above which the reverse is true.

The performance of SRMA has also been compared to that of polling techniques. Under a polling protocol, a master station asks the user stations one by one whether they have any messages to transmit. If a user station has information, it responds to the poll by sending all of the messages in its buffer. If a station has no information, it transmits a negative reply or no reply. For populations of more than 100 user stations, the delay performance of SRMA is shown in [Tobagi, 1974] to be superior to that of polling.

9.3 Simulation

The delay D_1 was computed using simulation results. Moreover, simulation studies were performed to verify the assumption that the arrival of successful requests to the scheduling station is Poisson [Tobagi, 1974].

9.4 Discussion

One advantage of the SRMA scheme is that in, contrast to some of the other packet techniques such as pure ALOHA and slotted ALOHA, SRMA can readily accommodate variable length or multi-packet messages in addition to fixed-length packets. Moreover, as

long as the ratio η of control information bits to message information bits is low, the maximum throughput achievable under SRMA is very high. However, as the number of user stations increases, so does η, since the amount of addressing information directly impacts the length of a request packet.

A serious limitation of SRMA for some applications lies in the instability of the request channel. Since it is operated in a random access mode, the request channel is vulnerable to saturation. Thus, for applications in which a minimum level of throughput must be guaranteed, SRMA may be unsuitable.

Both SRMA schemes were initially designed to operate over ground radio rather than a satellite channel. While the propagation delays associated with SRMA are minimal for a ground radio channel, these delays are substantial for a satellite channel. Recall that under SRMA, sending a message requires three separate transmissions. For a satellite operated under the SRMA protocol, this results in a propagation delay of approximately 0.75 s per message.

10.0 RESERVATION-ALOHA

10.1 Description

Reservation-ALOHA (or R-ALOHA) [Crowther, 1973] may be described as a TDMA system in which contention is used to initialize ownership of each time slot. Unlike *fixed* TDMA, however, R-ALOHA prohibits a station with no traffic from using channel capacity; a time slot "owned" by a station that no longer has traffic is automatically released and made available to other stations with traffic.

Under R-ALOHA, channel time is slotted and, as in other slotted schemes, a station must synchronize the start of a packet transmission with the beginning of a slot. Moreover, a frame structure is used. A requirement of the protocol is that each station receives not only its own, but *all* packet transmission over the channel.

The procotol works as follows. A station with a packet to transmit uses knowledge of channel activity in the previous frame to determine whether it may transmit in a particular slot in the current frame. Each of the M slots in the current frame is classified as "owned" or "empty". The Ith slot is owned by a given station if, during the previous frame, that station successfully transmitted a packet in the Ith slot. The Ith slot is classified as empty if, during the previous frame, either no packets were transmitted in that slot or two or more packets collided in that slot. A station may always transmit in a slot that it owns but is prohibited from transmitting in a slot owned by another station. Empty slots, however, are available to all stations on a contention basis. To prevent excessive collisions in the empty slots, a station may transmit in a given time slot with probability p, where p is less than one. Such a policy serves to control the transmission rate of new packets in the empty slots; moreover, it provides randomization of the retransmission delay for a previously collided packet.

If the communications channel is assumed to be a satellite channel, with the number of slots in a frame defined as the number of slots in one satellite round-trip delay, then

a user station takes action in the current time slot based on the transmission in the previous time slot which it has just received. See Figure 10.1 for an example of how R-ALOHA operates with a satellite channel.

10.2 Analysis

None.

10.3 Simulation

Preliminary simulation results for channel performance under R-ALOHA have been reported by Rettberg [Rettberg, 1973]. For the simulations, Rettberg assumes a 50 kilobit/second satellite channel and a Poisson input source which produces, with equal probability, two classes of traffic: single-packet and eight-packet messages.

For the simulations, the R-ALOHA protocol was specified in somewhat more detail than the above description and slightly modified. For example, in the simulations, the probability p that a station transmits in an empty slot is allowed to vary with the channel traffic; in light traffic, p is large, while p is reduced under heavy traffic. Note that this is a form of dynamic channel control. See [Rettberg, 1973] for details of additional modifications.

In Figure 10.2, the throughput-delay tradeoffs for single-packet messages are shown for a varying number of stations. Figure 10.3 displays the tradeoffs for eight-packet messages (the average delay shown in Figure 10.3 is for each packet of an eight-packet message; it is not the average delay for the total message). Note that in each figure, an increase in channel throughput is achieved at the expense of increased packet delay; in addition, the channel capacity of R-ALOHA approaches unity.

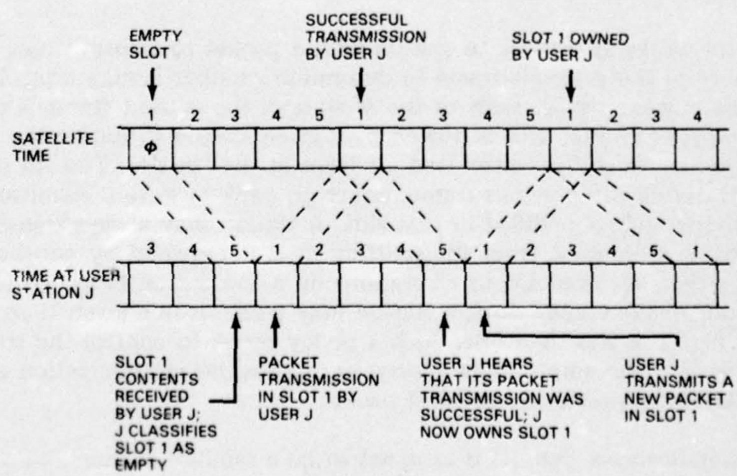


Fig. 10.1—Reservation—ALOHA: an example

10.4 Discussion

A major advantage of R-ALOHA is that it easily accommodates the entry of newly active stations to the communications system. Unlike a similar scheme, Binder's Round Robin (see Section 11), no slot is permanently assigned to a particular station; the ownership of slots changes along with the composition of the user population.

Although R-ALOHA was designed to operate with a satellite channel, the protocol could also be implemented with a ground radio channel. In the latter case, each user station would be required to store one frame's worth of the most recent channel activity, while in the case of the satellite channel, the only information necessary is the contents of the most recently received time slot.

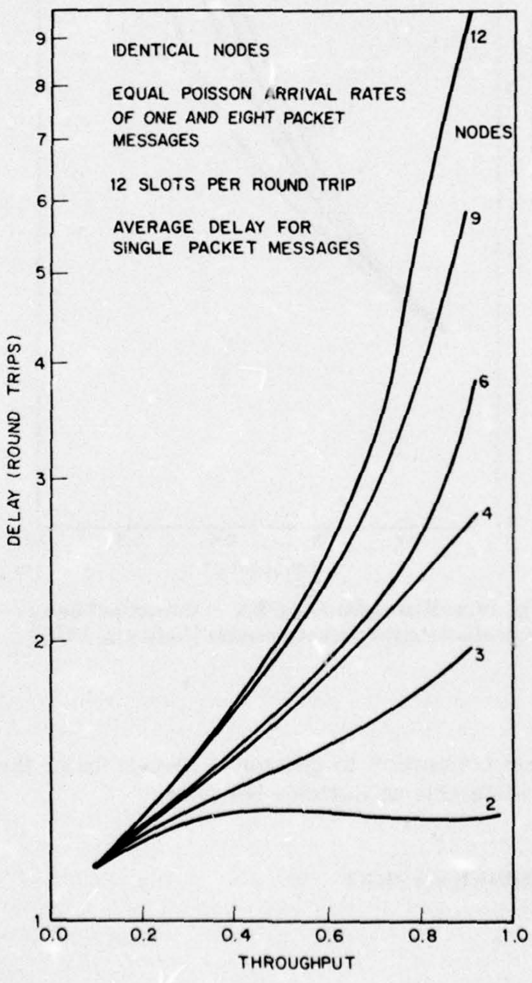


Fig. 10.2—Reservation-ALOHA: throughput-delay tradeoffs for single packet messages [Rettberg, 1973]

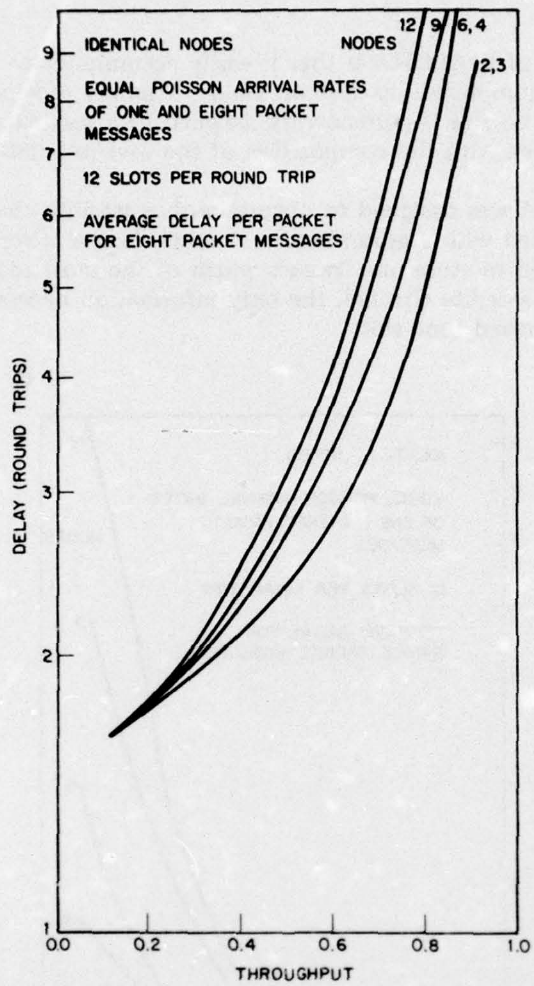


Fig. 10.3—Reservation-ALOHA: throughput-delay tradeoffs for eight-packet messages [Rettberg, 1973]

Since R-ALOHA uses contention to determine ownership of the time slots in a frame, the technique is vulnerable to unstable behavior.

11.0 ROUND ROBIN RESERVATION

11.1 Description

The round robin (RR) reservation technique, proposed and analyzed by Binder [Binder, 1975], is based on the use of a fixed TDMA structure coupled with a dynamic assignment system.

For this technique, channel time is divided into slots of fixed-length T seconds, where T includes the packet transmission time plus a synchronization and guard time. The slots are embedded in a frame structure, where each frame contains a fixed number of slots. Let T_{RMAX} represent the satellite round-trip propagation delay in seconds, and let N represent the number of user stations sharing the channel. Then T_F, the frame time, is fixed and must satisfy the condition

$$T_F \geqslant \mathrm{Max}\{NT, T_{RMAX} + T\}.$$

Thus, a frame is large enough to include one slot for each user. Moreover, since T_F is at least as large as $T_{RMAX} + T$, a slot will always be completely received by all users before it is time to transmit into that slot in the following frame.

Let K represent the number of slots in a frame ($K \geqslant N$). Then, the fixed TDMA component of this protocol consists of assigning slot 1 to user 1, slot 2 to user 2, . . . , slot N to user N. Fixed assignments are not made for slots $N + 1$, $N + 2$, . . . , K. This leads to the first rule of the protocol:

*R*1: A user may send in its fixed assignment slot at any time.

At any given time, a user station is *either active (it has traffic to send) or inactive* (it has no traffic to send). Those slots in a frame which belong to inactive users as well as the unowned slots in each frame form a pool of slots which are available for dynamic assignment to each currently active user. These are available in addition to an active user's owned slot in the frame. However, an active user must always use its "owned" slot in a frame before it can acquire the dynamically controlled slots.

The problem of making short-term allocations of the dynamically controlled slots with the long satellite propagation delay ($\approx$ 0.25 s) is solved by the use of a distributed queue similar to that proposed by Roberts in his reservation scheme [Roberts, 1973] (see Section 8). The distributed queue is managed as follows. Once every frame, each active station sends reservation information which reflects the state of its local queue of packets. This information is sent as part of the overhead of the data packet in the active user's owned slot. All users (both active and inactive) receive and store the information in a Channel Queue Table (CQT) which contains one entry for each user in the system; there is a zero entry in the CQT for each currently inactive user. A pointer is kept by all users and is used to keep track of the active user who received the last dynamically assignable slot. At each new dynamic assignment, this pointer is moved. In this way, the dynamically assignable slots are allocated one at a time among the active users, and no active user receives a second dynamically assignable slot until each other active user has received one.

When a previously inactive station wishes to send a packet, it uses *R*1 and transmits in its own slot, deliberately generating a conflict. One round trip later, the conflict is detected by all of the other users. To allow the newly active user to use its slot, the following rule is needed:

R2: A user may send in a dynamically assignable slot unless a conflict was received during that slot's last occurrence.

This guarantees channel access after one frame time to a previously inactive user, even though another user may be currently using his slot.

11.2 Analysis

None.

11.3 Simulation

The performance of this technique under varying traffic conditions was investigated by means of a simulation program [Binder, 1974]. Two classes of traffic were generated for each user: short messages consisting of single packets and long messages with eight packets per message. Poisson arrivals were assumed, with a different mean for each message class. Moreover, at each station, a priority queueing system was used; i.e., packets were always sent from a short message queue first and from a long message queue only after the short message queue was empty. When a packet transmission sustained a conflict, the packet involved was placed at the head of its queue.

For the simulations, a frame size of 12 slots was chosen; moreover, since the population size was 12 users, all of the slots were owned. Thus, dynamically assignable slots *were available only when one or more nodes were inactive*. The results of the simulations are shown in Figures 11.1-11.5 [Binder, 1974].

Figures 11.1 and 11.2 illustrate the throughput-delay performance of the RR technique for short message packets and long message packets, respectively. Packet delay, which is expressed in terms of satellite round trips (RT's), includes all queuing, transmission, and propagation times. The throughput rate shown in the figures is based on an input source containing both short and long messages. For comparison with RR, the performance of three other reservation schemes—TDMA, Reservation-ALOHA (Section 10), and Roberts' Reservation scheme (Section 8)—is also shown. The TDMA scheme is based on the use of a fixed assignment of one slot per user; it was obtained by inhibiting the dynamic assignment of the RR algorithm.

As shown in Figure 11.1, the RR scheme gives good delay performance for packets associated with short messages as the throughput rate increases above 0.4. The worst delay occurs for throughput of around 0.8 when the greatest number of conflicts occur due to intermittent activity at the user stations. As the throughput rate increases above 0.8, the long message queue at each station is rarely empty, no conflicts occur, and hence, packet delay approaches that of fixed TDMA.

Packet delay associated with long messages is shown in Figure 11.2. Again, the RR scheme performs very well when the throughput rate is high. However, the results for Roberts' Reservation scheme are not directly comparable, since they are based on average message delay, not packet delay. The Reservation-ALOHA scheme does better than RR under light loading conditions, i.e., when the throughput rate is less than 0.4.

Figure 11.3 shows the sensitivity to varying traffic mixes of average packet delay for both message classes. The mix is given by (X: Y), the ratio of short message arrivals to long message arrivals. For example, (8:1) is an equal packet mix, (0:1) is long messages

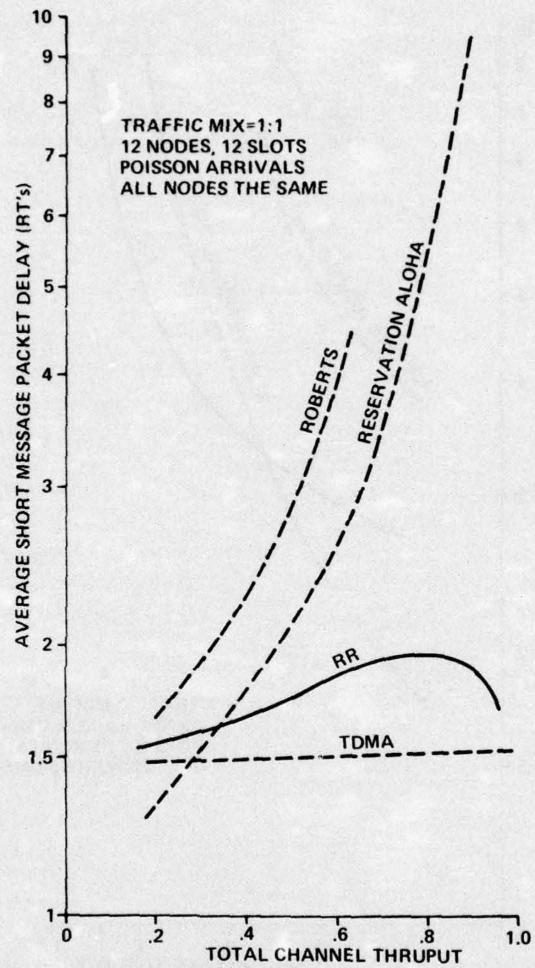


Fig. 11.1—Packet delay vs throughput for short messages [Binder, 1974, p. 14]

only, and (1:0) is short messages only. The total traffic is assumed to be equally distributed among all users. Long message packet delay for RR is relatively insensitive to variations in the traffic mix. However, the packet delay associated with short messages approaches that of fixed TDMA as the proportion of short messages increases. In particular, when the traffic consists of only short messages, the performance of RR is very similar to that of TDMA, indicating that gains due to reservations are being balanced out by conflicts in owned slots. On the other hand, when the traffic is composed of only long messages, RR performs better than TDMA, with much shorter delays for a given throughput.

The effect of an uneven traffic mix among users has also been considered, and the results are shown in Figures 11.4 and 11.5. In these figures, one user is considered a large user, while the others are small users. The input rate of each of the small users is the same, and each generates only short messages. The input from the large user consists only

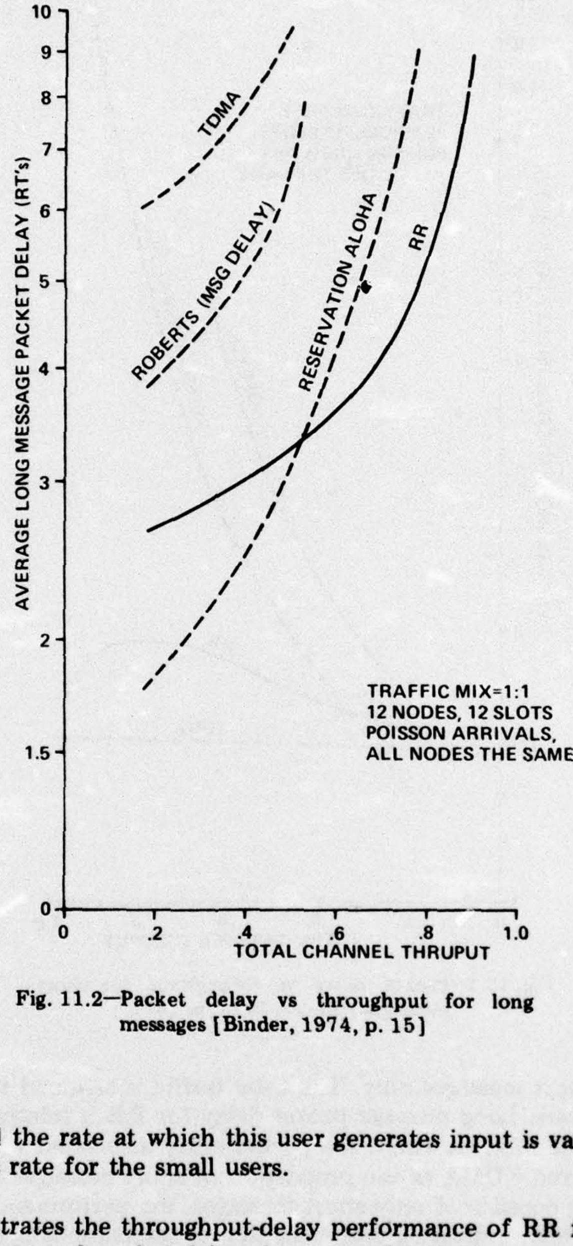


Fig. 11.2—Packet delay vs throughput for long messages [Binder, 1974, p. 15]

of long messages, and the rate at which this user generates input is varied. Let S_1 represent the total input rate for the small users.

Figure 11.4 illustrates the throughput-delay performance of RR for various values of S_1. The throughput rate and packet delay shown are for the large-user messages only. For comparison, the performance for a single user under fixed TDMA is also shown; assuming that the total number of users is N, a single user is assigned a fixed portion of the channel bandwidth, one-Nth, regardless of the input rates of the other users. At the other extreme is the curve labeled $S_1 = 0$, which shows the RR performance when there is no input from the small users; in this case, the total channel is available to the large user, and thus the packet delay is the sum of the average waiting time for a single-server queue plus

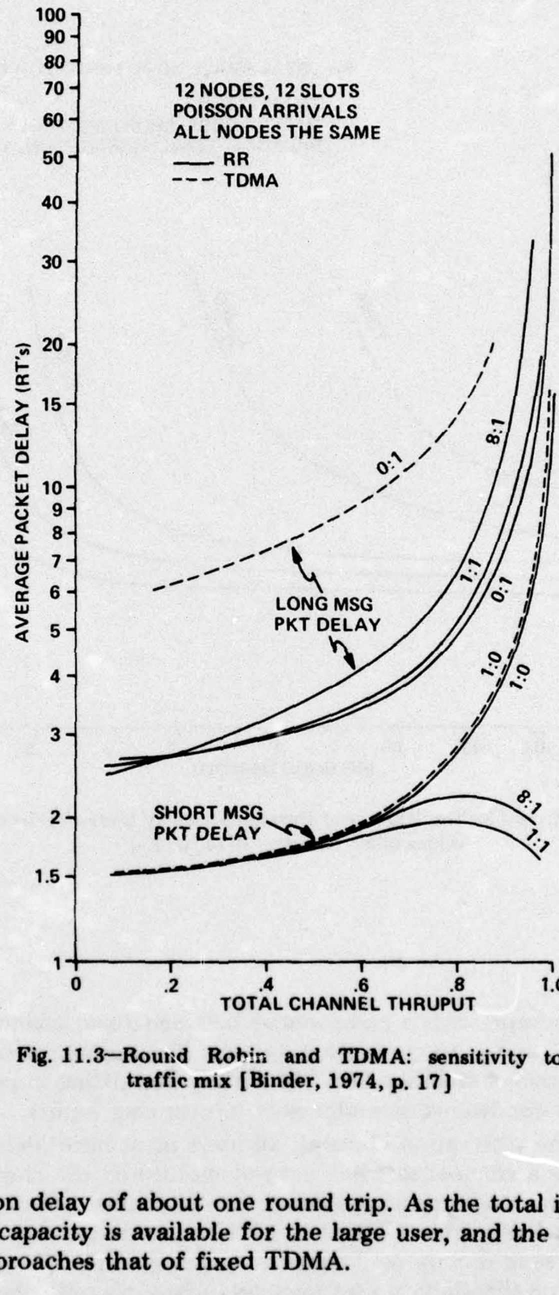


Fig. 11.3—Round Robin and TDMA: sensitivity to traffic mix [Binder, 1974, p. 17]

a constant reservation delay of about one round trip. As the total input from the small users increases, less capacity is available for the large user, and the throughput-delay performance of RR approaches that of fixed TDMA.

Another result of interest is the effect of the large user traffic on the packet delay for small users. Figure 11.5 shows the packet delay for small users as a function of the large-user throughput. Note that for each value of S_1, the packet delay for small users increases as the throughput rate of the large user increases. However, the total increase in packet delay is only about the one round trip required to gain access to the user's owned slot.

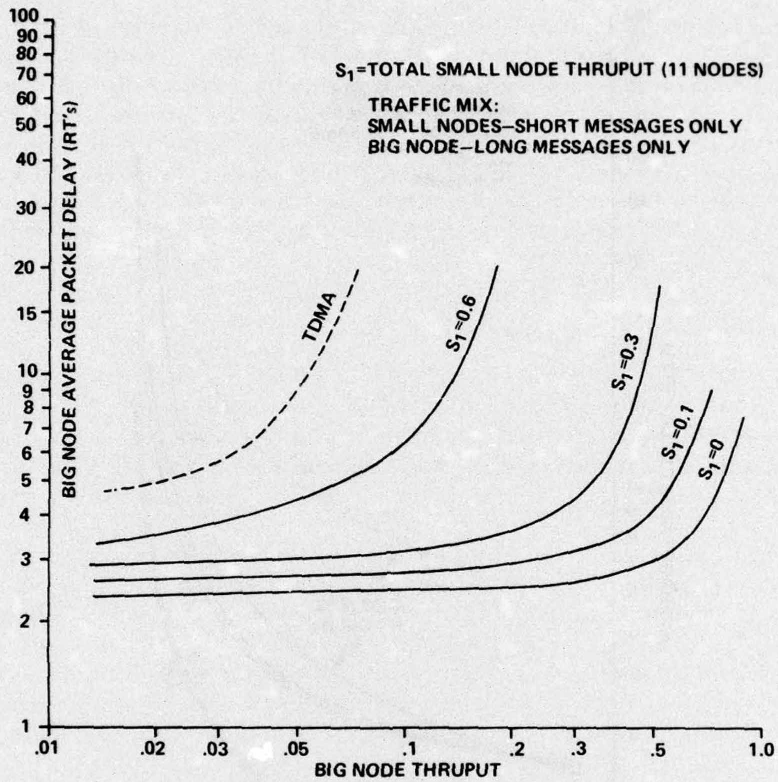


Fig. 11.4—Round Robin: larger user throughput-delay tradeoffs for different values of S_1 [Binder, 1974, p. 23]

11.4 Discussion

The RR technique represents a compromise between fixed channel management schemes such as TDMA and pure contention systems like random access ALOHA. The RR approach retains the channel stability of TDMA while permitting improved channel utilization under light load conditions or under nonuniform user inputs.

To make use of the reservation concept, all users must have identical copies of the CQT. A user who owns a channel slot but has not monitored the channel activity must not only achieve slot and frame synchronization but also must obtain a copy of the CQT. One means of handling this problem [Binder, 1975] is to designate a master station and require that station to send a copy of its CQT at the beginning of every frame, just before the data slots. Using this data, a user may resynchronize with the other channel users. However, there must be a preestablished protocol which allows a backup master station to take over if there is a failure of the current master station.

Errors in received reservation information may also force a user to resynchronize. To reduce these types of errors, the technique suggested by Roberts [Roberts, 1973] of sending three separate parity-checked copies of the reservation information may be used.

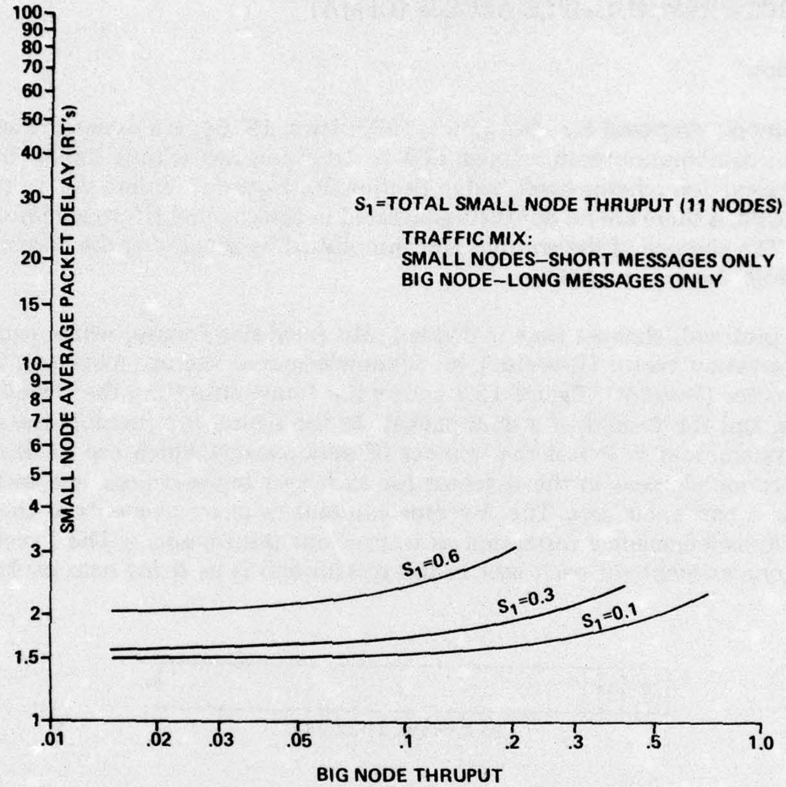


Fig. 11.5—Round Robin: small user packet delay for one large user [Binder, 1974, p. 24]

This results in additional overhead for the reservations, but may be insignificant given a high bandwidth channel and/or a limited number of users.

The analysis described above was based on the assumption that there is sufficient channel bandwidth so that the limiting factor on frame time is the round trip propagation delay. For large numbers of nodes ($\approx$ 100), this implies a wide bandwidth channel, Reducing the packet size for narrow bandwidth channels increases total packet overhead per frame and thus decreases channel utilization.

The round robin protocol described above is based on the assumption of an unchanging user population. For military applications in which the composition of the user population in a given area may be quite dynamic, this protocol provides no explicit means of modifying slot ownership to reflect changes in the user population (fixed TDMA, however, also has this problem).

12.0 CONFLICT-FREE MULTIPLE ACCESS (CFMA)

12.1 Description

This protocol, proposed by Hwa [Hwa, 1975; Hwa, 1976], is a dynamic assignment system used in combination with a fixed TDMA structure, and is thus similar to the round robin reservation scheme described in Section 10. However, unlike the round robin scheme, with CFMA there are no conflicts generated in the channel (if we assume an errorless channel). The absence of the conflict is accomplished by separating the reservation information from the data packets.

For this protocol, channel time is divided into fixed-size frames, where each frame contains a reservation vector (R-vector), an acknowledgment vector (A-vector), and an information vector (I-vector). Figure 12.1 shows the frame structure, the structure of each of the vectors, and the format of a data packet. In the figure, n represents the number of users in the system, and $m \geqslant n$ is the number of data packets which can be sent in each frame. There is one element in the R-vector for each user in the system, and each element is dedicated to a particular user. The A-vector contains as many elements as the I-vector and carries acknowledgment information as to previous transmissions. The I-vector contains at least one element for each user in the system and is used for data packet transmission.

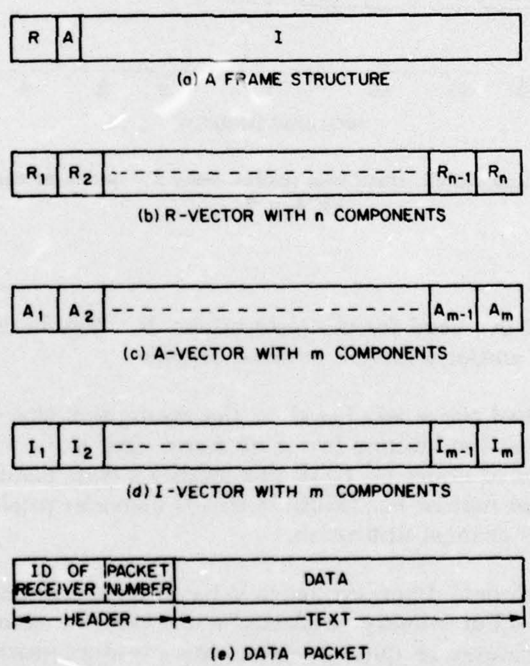


Fig. 12.1—Conflict-Free Multiple Access: frame structure and its components [Hwa, 1976, p. 4]. (Copyright 1976, Computer Science Department, University of Sydney, Australia, used by permission of the author.)

The elements in the I-vector in each frame are allocated to users on the basis of (a) the reservation information in the R-vector for that frame and (b) a fixed priority structure for each element of the I-vector ($I_1, I_2, \ldots. I_n$). The priority structure for element I_i is

$$U_i, U_{i-1}, \ldots, U_1, U_2, \ldots, U_{i+2}, U_{i+1} ;$$

i.e., user U_i has the highest priority for element I_i, user U_{i-1} has the second highest priority, etc.

The allocation algorithm works in the following manner. In frame F^t, each user sends, in its R^t-vector element, the number of data packets it wishes to send in the next frame, F^{t+1}. Any user with packets to transmit automatically uses the element(s) in I^{t+1}, for which it has the highest priority relative to the other users with traffic. Note that each user is always guaranteed at least one I-vector element.

As an example, consider a four-user system, where the I-vector has four elements. If the R-vector for frame F^t is

$$(1, 2, 0, 2)$$

then the I^{t+1}-vector elements in frame F^{t+1} are allocated to the users as follows:

$$(U_1, U_2, U_2, U_4) .$$

U_1 uses its highest priority I-vector element, I_1; U_2 uses its highest priority I-vector element, I_2, and U_2 also uses I_3. I_3 is available to U_2, because this user has the second highest priority for I_3, and the user with the highest priority for that I-vector element, U_3, has no traffic to send. Finally, U_4 uses its highest priority element, I_4.

Each user maintains information as to which I-vector elements it used in a given frame. This information, in conjunction with the acknowledgment information in the A-vector, is used to determine successful receipt of a data packet. Note that in CFMA, an unsuccessful packet transmission is due to noise on the channel or receiver problems, not to conflicts or collisions with other data packets in the I-vector.

12.2 Analysis

None.

12.3 Simulation

A simulation model was used to investigate the performance of this protocol and the results are reported by Hwa [Hwa, 1976]. The performance measures used are not compatible with those previously described and thus are not summarized here. However, the channel was shown to be highly stable and to provide channel utilization superior to fixed TDMA under light load conditions. When each user has a heavy traffic load, performance is somewhat worse than fixed TDMA due to the overhead of reservation information. Unfortunately, a comparison of the performance of this technique to that of the round robin reservation scheme (Section 11) is not available.

12.4 Discussion

As indicated above, CFMA is similar to the round robin (RR) protocol. (See Section 11 for details.) In the RR protocol, an inactive user whose owned slot is in use regains control of that slot by generating a conflict. Since, in CFMA, an inactive user may become active at any time by transmitting a nonzero value in its R-vector element, the need to generate a conflict disappears. The separation of reservation and data information in the conflict-free scheme eliminates a problem associated with RR, that of providing a user with a copy of the current Channel Queue Table (CQT), since the status of the channel and the allocation of data for frame F^{t+1} is fully contained in the R-vector for F^t.

One possible disadvantage of CFMA is that the fixed priority structure can result in one user having a blocking effect on another. For example, consider a four-user system similar to the one described above. Suppose that the R-vector for F^t is (1, 2, 0, 2). Then, the allocation for I-vector in F^{t+1} is (U_1, U_2, U_2, U_4). Suppose that at F^{t+1} the R-vector is (1, 2, 0, 3). Then, the allocation for F^{t+2} is also (U_1, U_2, U_2, U_4). In each frame, there is one empty I-vector element that belongs to an inactive user, U_3. Even though both U_2 and U_4 have excess traffic, U_2 always receives the empty vector element. The behavior of round robin in the example is different; that protocol would provide an allocation first to U_2, then to U_4, etc., and is thus fairly insensitive to the distribution of traffic among the users. In the conflict-free system, overall performance may be improved by adjusting priorities as a function of the expected traffic in neighboring slots.

13.0 SUMMARY AND CONCLUSIONS

We reviewed a variety of packet switching techniques designed to allocate dynamically a broadcast channel among multiple users. Under time-varying loads, each of these techniques does in fact make more efficient use of the channel than do fixed allocation schemes such as TDMA or FDMA. The packet techniques exploit the broadcast nature of the communications channel; they are applicable both to ground radio channels and to satellite channels, with the exception of CSMA, which is primarily suited to a ground radio channel.

At this stage of our study, it is possible to state several general conclusions about the techniques reviewed in this report:

(a) The ALOHA random access techniques, i.e., classical ALOHA, slotted ALOHA, and CSMA, have characteristics which could prove advantageous for Naval applications.

The relative insensitivity of these techniques to additions or losses in the user population is one advantage, while another is the degree of flexibility that the ALOHA schemes provide with respect to changing user communications requirements. Unfortunately, the ALOHA schemes have a serious shortcoming for military applications: they are dangerously vulnerable to unstable behavior and channel saturation. A sudden surge in the input load or stochastic variations in the channel traffic can cause the throughput of an ALOHA channel to deteriorate to zero with virtually no chance for recovery. This is unacceptable in most military applications where, in a crisis situation, the load on the communications system tends to increase and, at the same time, a minimum level of communications capability must be guaranteed. Also, the ALOHA schemes have been designed to handle traffic composed primarily of single-packet messages. Thus, these schemes have only limited relevance to many Navy applications, where message length tends to be variable. Note that, whereas the ALOHA schemes by themselves may be unacceptable for most military applications, these schemes may prove attractive as part of other multiplexing techniques; e.g., an ALOHA scheme might be used to transmit reservation requests, as in Roberts' reservation or in SRMA.

(b) Although dynamic control procedures were described in the context of the slotted ALOHA technique, these procedures may be used with other ALOHA schemes (pure ALOHA, CSMA) to reduce the risk of channel saturation. These procedures were however, designed to handle *short-term* fluctuations in the traffic; *long-term* increases in the input load will require additional control measures.

(c) All of the techniques, except SRMA, have been designed to operate with a distributed control mechanism. Thus, the systems are not vulnerable to the loss of a master control station, a clear advantage for military applications.

(d) In many of the traffic models used for the analysis of broadcast packet switching, it is assumed that stations do not generate new traffic whenever a packet from that station is blocked (stations cannot queue locally generated traffic). This assumption is unlikely to be valid for Naval applications.

(e) The reservation techniques described in the report were designed to handle multi-packet messages. If, as we suspect, a significant fraction of Navy traffic consists of multi-packet messages rather than single-packet messages, the reservation schemes rather than the ALOHA schemes would appear to be most relevant to the Naval application. Moreover, two of the reservation schemes, the Round Robin scheme and Conflict-Free ALOHA, are stable and thus guarantee a minimum level of communications capability under heavy traffic conditions. Hence, these two schemes as well as the other reservation schemes are the most promising candidates.

In general, the performance analyses of the various broadcast packet switching techniques assumed interactive computer communications applications. Both the user and the traffic characteristics of the analytic models reflect these applications, which differ substantially from likely Naval applications. Hence, it is necessary to estimate the performance of broadcast packet-switching techniques (or an appropriate subset of these techniques) with analytic models and performance constraints that more accurately reflect Naval applications. An example of a more appropriate traffic model assumption is the ability of all Naval users to queue locally generated traffic; a more appropriate performance constraint is that delay requirements for certain Naval systems are not nearly

as stringent as those for interactive computer systems. Broadcast packet switching techniques must also be evaluated with respect to such Naval requirements as encryption and privacy, countermeasures, survivability, etc. The suitability of these techniques for use in Naval applications will be the subject of a subsequent report.

REFERENCES

[Abramson, 1970] N. Abramson, "The ALOHA System—Another Alternative for Computer Communications," U. of Hawaii, ALOHA System Technical Report B70-1, Apr. 1970.

[Abramson, 1973a] N. Abramson, "Packet Switching with Satellites," National Computer Conference, New York, 42:695-702 (June 4-8, 1973).

[Abramson, 1973b] N. Abramson, "The ALOHA System," in *Computer-Communications Networks*, Norman Abramson and Franklin F. Kuo, Editors, Prentice-Hall, Englewood Cliffs, N.J. 1973.

[Binder, 1974] R. Binder, "A Dynamic Packet Switching System for Satellite Broadcast Channels," U. of Hawaii, ALOHA System Technical Report B74-5, Aug. 1974.

[Binder, 1975a] R. Binder, "A Dynamic Packet Switching System for Satellite Broadcast Channels," International Conference on Communications, Conference Record, III, San Francisco, Calif. (June, 1975).

[Binder, 1975b] R. Binder, *et. al.*, "ALOHA Packet Broadcasting—A Retrospect," U. of Hawaii, ALOHA System Technical Report B75-4, Jan. 1975.

[Bortels, 1970] W.H. Bortels, "Simulation of Interference of Packets in the ALOHA Time-Sharing System," U. of Hawaii, ALOHA System Technical Report B70-2, Mar. 1970.

[Carleial and Hellman, 1975] A.B. Carleial, and M.E. Hellman, "Bistable Behavior of ALOHA-Type Systems," IEEE Trans. Com., COM-23, 4:401-409 (Apr. 1975).

[Crowther, 1973] W. Crowther, *et. al.*, "A System for Broadcast Communications: Reservation ALOHA," Proceedings of the 6th Hawaii International Conference on Systems Sciences, January 1973.

[Ferguson, 1975] M.J. Ferguson, "A Study of Unslotted ALOHA with Arbitrary Message Lengths," 4th Data Communications Symposium, Quebec City, Oct. 7-9, 1975.

[Hwa, 1975] H.R. Hwa, "A Conflict-Free ALOHA System," Sydney/ALOHA Working Paper 3, Computer Science Department, University of Sydney, Sydney, Australia Jan. 1975.

[Hwa, 1976] H.R. Hwa, "A Simulation Study of Packet Switching in Conflict-Free Multi-Access Broadcast Data Communications Systems," Sydney/ALOHA Working Paper 7, Computer Science Department, University of Sydney, Sydney, Australia Jan. 1976.

[Kleinrock, 1976] L. Kleinrock, *Queueing Systems, Volume 2: Computer Applications*, John Wiley and Sons, New York, 1976.

[Kleinrock and Lam, 1973] L. Kleinrock, and S.S. Lam, "Packet Switching in a Slotted Satellite Channel," National Computer Conference, New York, 42, 703-710 (June 4-8, 1973).

[Kleinrock and Lam, 1975a] L. Kleinrock, and S.S. Lam, "Dynamic Control Schemes for a Packet Switched Multi-Access Broadcast Channel," National Computer Conference, Anaheim, Calif. pp 143-153, May 19-22, 1975.

[Kleinrock and Lam, 1975b] L. Kleinrock, and S.S. Lam, "Packet Switching in a Multi-Access Broadcast Channel: Performance Evaluation," IEEE Trans. Comm. COM-23, 4:410-422 (Apr. 1975).

[Kleinrock and Tobagi, 1975a] L. Kleinrock, and F. Tobagi, "Random Access Techniques for Data Transmission Over Packet Switched Radio Channels," National Computer Conference, 1975.

[Kleinrock and Tobagi, 1975b] L. Kleinrock and F.A. Tobagi, "Packet Switching in Radio Channels: Part 1—Carrier Sense Multiple-Access Modes and Their Throughput Delay Characteristics," IEEE Trans. Comm. COM-23, 12:1400-1416 (Dec. 1975).

[Kobayashi, 1976] H. Kobayshi, *et. al.*, "An Approximate Method for Design and Analysis of an ALOHA System," U. of Hawaii, ALOHA System Technical Report B76-4, Mar. 1976.

[Konheim, 1974] A.G. Konheim and B. Meister, "Waiting Lines and Times in a System with Polling," J. Ass. Computer Mach. 21(3):470-490 (July 1974).

[Lam, 1974] S.S. Lam, "Packet Switching in a Multi-Access Broadcast Channel with Application to Satellite Communications in a Computer Network," U. of California, Los Angeles, Department of Computer Science, Ph.D. Dissertation, Mar. 1974.

[Lam and Kleinrock, 1975] S.S. Lam and L. Kleinrock "Packet Switching in a Multi-Access Broadcast Channel: Dynamic Control Procedures," IEEE Trans. Comm. COM-23, 9:891-904 (Sept. 1975).

[Metcalfe, 1973] R.M. Metcalfe, "Packet Communication," Project MAC, Massachusetts Institute of Technology, Report MAC TR-114, Cambridge, Mass. Dec. 1973.

[Rettberg, 1973] R. Rettberg, "Preliminary Simulation Results for Reservation ALOHA," ARPA Network Information Center, Stanford Research Institute, Menlo Park, Calif., Note 43, NIC Doc. 16086, May 1973.

[Roberts, 1972] L.G. Roberts, "ALOHA Packet System With and Without Slots and Capture," ARPA Network Information Center, Stanford Research Institute, Menlo Park, Calif., ASS Note 8, NIC Doc. 11290, June 1972.

[Roberts, 1973] L.G. Roberts, "Dynamic Allocation of Satellite Capacity Through Packet Reservation," National Computer Conference, New York, N.Y. 42:711-716, June 4-8, 1973.

[Tobagi, 1974] F. Tobagi, "Random Access Techniques for Data Transmission Over Packet Switched Radio Networks," U. of California, Los Angeles, School of Engineering and Applied Science, UCLA-ENGR 7499, Dec. 1974.

[Tobagi and Kleinrock, 1975] F.A. Tobagi and L. Kleinrock "Packet Switching in Radio Channels: Part II—The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy Tone Solution," IEEE Trans. Comm. COM-23, 12:1417-1433 (Dec. 1975).